

Db 61 ATGCGGCTGGGTATTCCGACCGAGACCAAAACAGCAATTCCGGGTGGCCATCACCCG 120  
 QY 121 GCCGGGCTCGGGAACATAACCCCTGCTGGCCATGAGGTGCTCATCCAGGAGTGC CGA 180  
 Db 121 GCCGGGCTCGGGAACATAACCCCTGCTGGCCATGAGGTGCTCATCCAGGAGTGC CGA 180  
 QY 181 GAGGCTCGGCTATACACCGAGCGGATTTCAAGGCGGACAGCCGCACTGCTCGGACC 240  
 Db 181 GAGGCTCGGCTATACACCGAGCGGATTTCAAGGCGGACAGCCGCACTGCTCGGACC 240  
 QY 241 GCCGACAGGTGTGGGCGGACGCTGATTTATGCTCAAGGCTCAAAAGACGATAGCGCG 300  
 Db 241 GCCGACAGGTGTGGGCGGACGCTGATTTATGCTCAAGGCTCAAAAGACGATAGCGCG 300  
 QY 301 GAATACGCGCGCTGCGACACGCGGAGATCTTGTACGTTTCTTGCATTTGGCGCGTCA 360  
 Db 301 GAATACGCGCGCTGCGACACGCGGAGATCTTGTACGTTTCTTGCATTTGGCGCGTCA 360  
 QY 361 CGTCTTGCACCGATGGTGTGTTGATTCGGGACACGATGCTTACGATAGAGACCGTC 420  
 Db 361 CGTCTTGCACCGATGGTGTGTTGATTCGGGACACGATGCTTACGATAGAGACCGTC 420  
 QY 421 CAGACCGCGACGCGGACGACTACCCCTGCTTGGCCGATC AGCGAAGTCCGCGGTGACATC 480  
 Db 421 CAGACCGCGACGCGGACGACTACCCCTGCTTGGCCGATC AGCGAAGTCCGCGGTGACATC 480  
 QY 481 GCGCGCGAGTTGGCGCTTACCACTGATGCGAACCAGGCGGCGCGGTGCTGATG 540  
 Db 481 GCGCGCGAGTTGGCGCTTACCACTGATGCGAACCAGGCGGCGCGGTGCTGATG 540  
 QY 541 GCGGCGGTGCGCGGCTGCGAACCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
 Db 541 GCGGCGGTGCGCGGCTGCGAACCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
 QY 601 TACAACGACGCGCGATCGGCAACGCGATGAGCGCGGCGGCGGCGGCGGCGGCGGCGG 660  
 Db 601 TACAACGACGCGCGATCGGCAACGCGATGAGCGCGGCGGCGGCGGCGGCGGCGGCGG 660  
 QY 661 ATGCAACAACTTGGGCAACTGAGCGCGGATTCGCGCGCGGATCACAACCTGCTACTCA 720  
 Db 661 ATGCAACAACTTGGGCAACTGAGCGCGGATTCGCGCGCGGATCACAACCTGCTACTCA 720  
 QY 721 TCGGCTTACGAGCTGAGGAGTGGCGTCAAAAGTCCGACCTGCTGATGGGCGGCTCTG 780  
 Db 721 TCGGCTTACGAGCTGAGGAGTGGCGTCAAAAGTCCGACCTGCTGATGGGCGGCTCTG 780  
 QY 781 GTGCGAGGCGGCAAGGACCAATTAAGTCTGCAATTCACTCGGCGATATGAAACA 840  
 Db 781 GTGCGAGGCGGCAAGGACCAATTAAGTCTGCAATTCACTCGGCGATATGAAACA 840  
 QY 841 GTGCGGCTACTGATGATATAGCCATGAGGAGGCGGCTGTTGAAAGGCTCAGACCG 900  
 Db 841 GTGCGGCTACTGATGATATAGCCATGAGGAGGCGGCTGTTGAAAGGCTCAGACCG 900  
 QY 901 ACACCTTACGACACCGGAGCTTGGCGTGCAGACACGCTGTTTACTGCTGCGGCAAC 960  
 Db 901 ACACCTTACGACACCGGAGCTTGGCGTGCAGACACGCTGTTTACTGCTGCGGCAAC 960  
 QY 961 ATGCGCGCTGCTGAGGAGGAGCTGCACTGAGCGGCTGACCAACGCGACGATGCGGAT 1020  
 Db 961 ATGCGCGCTGCTGAGGAGGAGCTGCACTGAGCGGCTGACCAACGCGACGATGCGGAT 1020  
 QY 1021 GTGCTGAGGCTTGGCGGACATGAGGAGGCGGCTGCGGCTGGAATCGGCGACTAGCC 1080  
 Db 1021 GTGCTGAGGCTTGGCGGACATGAGGAGGCGGCTGCGGCTGGAATCGGCGACTAGCC 1080  
 QY 1081 AAAGCTTTTTCAGACGAGGAGGCGGCTTACTGTCGAAACGCGTGGCCACGACCTGGGG 1140  
 Db 1081 AAAGCTTTTTCAGACGAGGAGGCGGCTTACTGTCGAAACGCGTGGCCACGACCTGGGG 1140  
 QY 1141 GTGCGGCTTACCGAGCGCGGCGGCGGCTGCTGAGCTGATC TGGCGGCTGCTTACGCGGAG 1200  
 Db 1141 GTGCGGCTTACCGAGCGCGGCGGCGGCTGCTGAGCTGATC TGGCGGCTGCTTACGCGGAG 1200

QY 1201 CACACGTCGGGAGTAAGGAGGAGGATGA 1228  
 Db 1201 CACACGTCGGGAGTAAGGAGGAGGATGA 1228  
 RESULT 3  
 A87607 1237 bp DNA PAT 22-JAN-2000  
 LOCUS A87607  
 DEFINITION Sequence 4 from Patent WO9836089.  
 ACCESSION A87607  
 VERSION A87607.1 GI:6736247  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1237)  
 FLOHE, L. and Singh, M.  
 TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source? 1..1237  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 236 a 394 c 386 g 221 t  
 ORIGIN  
 Query Match 100.0%; Score 1228; DA 5; Length 1237;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-165;  
 Matches 1228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATCTTGAGATTATATCTTCTTCTATCTGAAAGCTACAGTATCGAAGGAGGTAATC 60  
 Db 1 ATCTTGAGATTATATGAACTTTCTTCTATCTGAAAGCTACAGTATCGAAGGAGGTAATC 60  
 QY 61 ATGCGGCTCGGATATTCGACCGAGACCAAAACAGCAATTCGGGTGGCCATCACCCG 120  
 Db 61 ATGCGGCTCGGATATTCGACCGAGACCAAAACAGCAATTCGGGTGGCCATCACCCG 120  
 QY 121 GCGGCGTGGCGGAGATCAACCCGCTGCTGCGATGAGGTGCTCATCCAGGAGGTGCCGA 180  
 Db 121 GCGGCGTGGCGGAGATCAACCCGCTGCTGCGATGAGGTGCTCATCCAGGAGGTGCCGA 180  
 QY 181 GAGGCTCGGCTATACACCGCGGATTCAGGCGGAGCGGCGGCGGCGGCGGCGGCGG 240  
 Db 181 GAGGCTCGGCTATACACCGCGGATTCAGGCGGAGCGGCGGCGGCGGCGGCGGCGG 240  
 QY 241 GCCGACAGGTGTGGGCGGAGGCTGATTTATGCTCAAGGCTCAAAAGACCGATAGCGCG 300  
 Db 241 GCCGACAGGTGTGGGCGGAGGCTGATTTATGCTCAAGGCTCAAAAGACCGATAGCGCG 300  
 QY 301 GAATACGCGCGCTGCGACACGCGGAGATCTTGTACGTTTCTTGCATTTGGCGCGTCA 360  
 Db 301 GAATACGCGCGCTGCGACACGCGGAGATCTTGTACGTTTCTTGCATTTGGCGCGTCA 360  
 QY 361 CGTCTTGCACCGATGGTGTGTTGATTCGGGACACGATGCTTACGATAGAGACCGTC 420  
 Db 361 CGTCTTGCACCGATGGTGTGTTGATTCGGGACACGATGCTTACGATAGAGACCGTC 420  
 QY 421 CAGACCGCGGACGCGGAGTACCCCTGCTTGGCCGATCAGCGAAGTCCGCGGTGACATC 480  
 Db 421 CAGACCGCGGAGTACCCCTGCTTGGCCGATCAGCGAAGTCCGCGGTGACATC 480  
 QY 481 GCGCGCGAGTTGGCGCTTACCACTGATGCGAACCAGGCGGCGGCGGCGGCGGCGG 540  
 Db 481 GCGCGCGAGTTGGCGCTTACCACTGATGCGAACCAGGCGGCGGCGGCGGCGGCGG 540  
 QY 541 GCGGCGGTGCGCGGCTGCGAACCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
 Db 541 GCGGCGGTGCGCGGCTGCGAACCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600





```

Db      1081   AAGGTCCTTTGCAGCACCAGAAAGGGCGTACTGTCCGAACGGGATGGCACACGACACTGGGG    1140
Oy      1141   GTGCCGTTACCGAGACCCGCCAGCGTGCTGGCTCACTCTCGGCGGGTCGTATACCCGAG    1200
Db      1141   GTGCGGTTCACGAGCCCCGCGAGCTGCTGGCTCACTCTCGGCGGGTCGTATACCCGAG    1200
Oy      1201   CACAGCTGGGAGTAAGGAACAAGCATGA    1228
Db      1201   CACAGCTGGGAGTAAGGAACAAGCATGA    1228

RESULT      5
LOCUS       MTW002             56414 bp     DNA                    BCT              17-JUN-1998
DEFINITION  Mycobacterium tuberculosis H37RV complete genome; segment 122/162.
ACCESSION   AL008967 AL133456
VERSION     AL008967.1 GI:3261491
KEYWORDS
SOURCE      ORGANISM
            Mycobacterium tuberculosis.
            Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Acidimycetales; Corynebacteriaceae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1 (bases 1 to 56414)
AUTHORS    Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekla,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Felwell,T., Gentles,S.,
            Hamlin,N., Holtroyd,S., Hornsbury,T., Jagsels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Squires,K., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrrell,B.G.
TITLE       Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
JOURNAL    Nature 393 (6685), 537-544 (1998)
REMARK     Erratum:[published erratum appears in Nature 1998 Nov
            12;396(6707):1901]
            2 (bases 1 to 56414)
            Direct Submission
            Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT    On Jun 27, 1998 this sequence version replaced g1:2624256.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M.tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in Tbpase (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.

FEATURES             Location/Qualifiers
     source           1..56414
                     organism="Mycobacterium tuberculosis"
                     strain="H37RV"
                     db_xref="taxon:1773"
                     <1..>56414
                     /note="fragment designated v002. Does not represent a
                     physical clone"
                     <1..233
```

gene  
CDS  
*/organism="Mycobacterium tuberculosis"*  
*/strain="H37Rv"*  
*/db\_xref="taxon:1773"*  
*/clone="Y154"*  
complement(3..527)  
*/gene="recX"*  
complement(3..527)  
*/note="rv2736c." (MTV002\_01c), len: 174 aa. recx, similar*  
*to eg. REC\_X\_PSEAF\_P37860 regulatory protein recx from*  
*Pseudomonas aeruginosa (153 aa), fasta scores: opt: 161*  
*z-score: 257.2 E(): 3.6e-07, 30.7% identity in 137 aa*  
*overlap. Overlaps and extends CDS from overlapping cosmid*  
*MTCY154\_16c"*  
*/codon\_start=1*  
*/transl\_table=1*  
*/product="recx"*  
*/protein\_id="CAI15532.1"*  
*/db\_xref="GI:2624257"*  
*/translacion="MTVSCPPSPSTSEREQARALCLRLTARSRTRAELACQAKRGIT*  
*PEDIGNRYDLAAAGLVDDTDFAEOMVSRRANAKSKRALAEHLAKGVDDVVITTT*  
*VLDGDAEAERGRACKLVRARLRREVLIIDDTDEARVSRLVMALARIGQTACEYE*  
*VIAELAERRRRY"*  
complement(493..2855)  
*/gene="recA"*  
complement(493..2865)  
*/gene="recA"*  
*/note="rv2737c." (MTV002\_02c), len: 790 aa. recA,*  
*identical to RECA\_MYCTU\_P26345 recA protein (790 aa).*  
*Contains self-splicing protein element (intein) from 2294*  
*to 969 (C), similar to intein II from TR.E332317*  
*(EMBL:Y13030) DNA-directedDNA polymerase (EC 2.7.7.7) from*  
*Thermococcus sp. (1829 aa), fasta scores: opt: 81 z-score:*  
*235.2 E(): 6e-05, 24.6%identity in 183 aa overlap.*  
*Contains PS00017 ATP/GTP-binding site motif A (P-loop),*  
*PS00321 recA signature, and PS00881proteine splicing*  
*signature. See Davis et al, (1992) Cell171(2):201-210"*  
*/codon\_start=1*  
*/transl\_table=1*  
*/product="recA"*  
*/protein\_id="CAI15533.1"*  
*/db\_xref="GI:2624258"*  
*/translacion="MTOTPDREKALBELAVAOIEKSYGKGSWVLGDDEAROPISVTPT*  
*DIARKLGVDTSLSVSDPTEGOALEIADMLIRSGSLDIVDYSVALVPRALEGEEN*  
*GSHVGLOLRMSQALRKMTGALNNSGTVAIFINOLRDIGVFESPETTGKVKDK*  
*VASYNDRVERVELTKDGTNAVGNTRRVKKAKCAEGRIIDPTGTHREDVADK*  
*RERIHVAANKDGLHARPVSMFDOGTADVIGLILAGAIVATPDKRVLFLEYCM*  
*ADKLKGGDVAAQPRRDGFGDSAPFPADHARLLGYLIGSRPGWGKPIEFINVQ*  
*AIIIDVTRLAATLGCALHPGRGISLAHAHRGERGVADLCODAGITCKIMERTIP*  
*KFFEPRIADTYNCGLFLFCIESDGWSREQOTALRPQYTTFECLAHOIHILLREG*  
*GSTRVIDDYQRPSITLVNGRRIQSRKFVEVISGMNVTAFAESVPMKPRGALITI*  
*AIAPEARQRRGSSQAITYLAAEMTDVLANLTDRGVYAQAAAMIGVASDPDGSKQO*  
*LASRSRRORVOLADALDDCKTLHDMLAEELYSYIREVLPTRARRIDPLEEELHTI*  
*VEGVNVRHSPPFKOAEPFDILYKGISREGLIDMGVDOGILRSKGAMFTVEGSQLD*  
*OGKEVNRNLTVENADVADEILEKKIKEKIGICAVVTDDPSENDVLPAVDV"*  
complement(790..2112)  
*/gene="recA"*  
*/note="recA intein"*  
complement(790..807)  
*/gene="recA"*  
*/note="PS00881 Protein splicing signature"*  
complement(2197..2223)  
*/gene="recA"*  
*/note="ps00321 recA signature"*  
complement(2644..2667)  
*/gene="recA"*  
*/note="PS00017 ATP/GTP-binding site motif A (P-loop)"*  
complement(2873..2877)  
*/note="possible RBS upstream of Rv2737c"*  
complement(3247..3253)

```

CDS
/gene="RV2738c"
/complement(3247..3453)
/note="RV2738c"
/note="RV2738c, (MTV002.03c), len: 68 aa. Unknown, N-terminus is highly similar to the N-terminus of the upstream ORFMTV002.07c (78.4% identity in 37 aa overlap); also similar to ALO20958|SC4H8.5 Streptomyces coelicolor cosmid 4H8 (64aa) opt: 185 z-score: 283.5 E(): 2.9e-08; 39.7% identity in 63 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein RV2738c"
/protein_id="CA15534.1"
/db_xref="GI:2624260"
/db_xref="SPTREMBL:O33281"
/transl_table="MAGVRLTEPHERALHFGAAYGSSVLLDHYLTGFDGRSAQAI EDGVEPRVWRAICADFDVPHDRW"
/complement(3464..4630)
/gene="RV2739c"
/complement(3464..4630)
/note="RV2739c"
/note="RV2739c, (MTV002.04c), len: 388 aa. Probable transferase, similar to eg. TR:Q51560 (EMBL:U28170) pseudomonas aeruginosa thiamosyl transferase (426 aa); fasta scores: opt: 178 z-score: 226.3 E(): 1.9e-05; 25.9% identity in 425 aa overlap. Equivalent to Mycobacterium leprae protein MCB33.02c (392 aa); fasta scores: opt: 294723|MCB33.2 Mycobacterium leprae cc nid B33 opt: 2112 z-score: 2364.5 E(): 0; 80.9% identity in 388 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein RV2739c"
/protein_id="CA15535.1"
/db_xref="GI:2624261"
/db_xref="SPTREMBL:O33282"
/transl_table="MRVAVVAGPDPGSHSPALAIQORPRAADPTLTFTGVEMLEARN AGTDAVELDGLATDRDLDAARIRHRAAOMATLVNRLALPELVSVYTRACG MAELLGIPWELNPHPLPLPSKGLPPIGSLAGTGRGLRDTMTALGRSWRG LROARAVREIGLIPARDPGLRLIATLPALEVPDPDAVAVVGLPHEPPTRVIA IPAGPVPVAVPSTALGTAGLLEVALQSILPGTVPVSGSLVSLSGADLVPPV AVAGSGOELLTRADLVICGGGGMVKTLLAGPVMVVGCGQMFARVYRQGS AVLIRPTADALVAIVNEVLSPPREARARAAASVAAADPVYCHALALAC"
/4674..5123
/gene="RV2740"
/4674..5123
/gene="RV2740"
/note="RV2740"
/note="RV2740, (MTV002.05), len: 149 aa. unknown, similar to M. leprae hypothetical protein TR:Q49850 (EMBL:294723) MCB33.03 (48235_F3_140) (178 aa) fasta scores: opt: 498 z-score: 538.9 E(): 7.3e-23; 51.6% identity in 161 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein RV2740"
/protein_id="CA15536.1"
/db_xref="GI:2624262"
/db_xref="SPTREMBL:O33283"
/transl_table="MAELTETSPEPTTEAIRAVEAFNLQNEDEFTVDAACDIL VYENYGFIRIGRRTATLIRMOGRVGFVGVIRHIGAGAVLTERDALIIPLRV OFWVGVEVDFGRITLMDYFDYDMFKGLRGLHVALVPSLKATL"
/5355..6932
/gene="RV2741"
/5355..6932
/gene="RV2741"
/note="RV2741"
/note="RV2741, (MTV002.06), len: 525 aa; Member of M. tuberculosis PE_PGRS subfamily, similar to many eg.
Query Match 100.0% Score 1228; DB 1; Length 56414;
Best Local Similarity 100.0% Pred No. 1.8e-16;
Matches 1228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCTGCAGATTATGCAACTTCTTCATCTACTAGAGCTACAGTATGAGAGGGTAAATC 60

```



Query Match	99.9%	Score 1226.4	DB 5	Length 1236
Best Local Similarity	99.9%	Pred. No. 7.2e-169		
Matches 1227	Conservative	0	Mismatches 1	Indels 0
QY 1	ATCTTCGAGATTAAACGAACTTCTTCATCTAGAACGCTACAGTATCGAGAGGGTAAATC	60		
Db 1	ATCTTCGAGATTAAACGAACTTCTTCATCTAGAACGCTACAGTATCGAGAGGGTAAATC	60		
QY 61	ATGCCGCTCGGATTCCTCCGACCGAAGCCAAACAAACGATTCGCGGTGGCCATCACCCG	120		
Db 61	ATGCCGCTCGGATTCCTCCGACCGAAGCCAAACAAACGATTCGCGGTGGCCATCACCCG	120		
QY 121	GGCGGCTGCGCGAACTAACCCGTCGTGGCCATGAGGTCTCATCCAGGAGGTGCCGA	180		
Db 121	GGCGGCTGCGCGAACTAACCCGTCGTGGCCATGAGGTCTCATCCAGGAGGTGCCGA	180		
QY 181	GAGGGCTCGGCTATACCGACGCGGATTTCCAAAGCGGCGAGCGGCGACTGGTGGCACC	240		
Db 181	GAGGGCTCGGCTATACCGACGCGGATTTCCAAAGCGGCGAGCGGCGACTGGTGGCACC	240		
QY 241	GCCGACAGAGTGTGGGCGCAGCGTGATTTATGCTCAAGATCAAAACCGATAGCGCG	300		
Db 241	GCCGACAGAGTGTGGGCGCAGCGTGATTTATGCTCAAGATCAAAACCGATAGCGCG	300		
QY 301	GAAATAGCGCGCTGCGACAGCGGCGAGATCTTGTTCACGTTCTTGCAATTTGGCCGCTCA	360		
Db 301	GAAATAGCGCGCTGCGACAGCGGCGAGATCTTGTTCACGTTCTTGCAATTTGGCCGCTCA	360		
QY 361	GATGCTGACCGAGATCGTTGTTGATTCGCGACCGCAGTCATTTGCTCAAGAGACGCT	420		
Db 361	GATGCTGACCGAGATCGTTGTTGATTCGCGACCGCAGTCATTTGCTCAAGAGACGCT	420		
QY 421	CAGACCGCGCAGCGCGCACTACCCCTGCTTGCCTCCGATGAGCAAGTCCGCGCTCACTC	480		
Db 421	CAGACCGCGCAGCGCGCACTACCCCTGCTTGCCTCCGATGAGCAAGTCCGCGCTCACTC	480		
QY 481	GGCGCCAGTGTGGGCTTACCACTGATGCGAACCAGGGGGCGCGGTGTGCTGATG	540		
Db 481	GGCGCCAGTGTGGGCTTACCACTGATGCGAACCAGGGGGCGCGGTGTGCTGATG	540		
QY 541	GGCGGGTCCCGGCGCTCAGAACCGGCGAGCTGTGATTCGCGCGCGGCGCGCGCG	600		
Db 541	GGCGGGTCCCGGCGCTCAGAACCGGCGAGCTGTGATTCGCGCGCGGCGCGCGCG	600		
QY 601	TACACGCGACCGCGCATTCGCCCAACGGCATGAGGCGCGACTTACGGTTCTAGACATCAAC	660		
Db 601	TACACGCGACCGCGCATTCGCCCAACGGCATGAGGCGCGACTTACGGTTCTAGACATCAAC	660		
QY 661	ATCGACAACCTTCGGAACCTCGACGCGCGAGTTCTGGGCGCGGATTCACACTCCCTACTCA	720		
Db 661	ATCGACAACCTTCGGAACCTCGACGCGCGAGTTCTGGGCGCGGATTCACACTCCCTACTCA	720		
QY 721	TGCGCTACGAGCTCGAGGCTGCGTCAACGTCGCGCACTGGTGAATGGGCGGCTCTG	780		
Db 721	TGCGCTACGAGCTCGAGGCTGCGTCAACGTCGCGCACTGGTGAATGGGCGGCTCTG	780		
QY 781	GTCGCGAGGGGCGCAAGGCAACCAATTAAGTGTGCAATTCATTGTGCGGCGATATGAAACCA	840		
Db 781	GTCGCGAGGGGCGCAAGGCAACCAATTAAGTGTGCAATTCATTGTGCGGCGATATGAAACCA	840		
QY 841	GGTGCAGTACTGTGATATAGCCATCGACAGGCGGCTGTTCGAAGGCTCAGCAGCG	900		
Db 841	GGTGCAGTACTGTGATATAGCCATCGACAGGCGGCTGTTCGAAGGCTCAGCAGCG	900		
QY 901	ACCACCTAGACCAACCGAGAGTTCGCGGTGACAGACAGCTGTTTACTCGTGGCGCAAC	960		
Db 901	ACCACCTAGACCAACCGAGAGTTCGCGGTGACAGACAGCTGTTTACTCGTGGCGCAAC	960		
QY 961	ATGCGCGGCTCGGTGCGAGAGAGTGCAGCTACGCGCTGACAAACGCGAGATGCCGTAT	1020		
Db 961	ATGCGCGGCTCGGTGCGAGAGAGTGCAGCTACGCGCTGACAAACGCGAGATGCCGTAT	1020		
QY 1021	GTGCTGAGCTTCGCGACCATGCTGCGGCGGCGGCTGCGGCTGGAATCCGGCACTAGCC	1080		

Db	1021	GTGCTCGAGCTTGCCGACCATGCTGGCGGGCGGTGCGGATCGAATCGGCACTAAC	1080
QY	1081	AAAGTCTTTTGAGCGCAGCAGAGGGGCGTTACTGTGCCAAGCGGTG6CCACCGAGCTGG66	1140
Db	1081	AAAGTCTTTTGAGCGCAGCAGAGGGGCGTTACTGTGCCAAGCGGTG6CCACCGAGCTGG66	1140
QY	1141	GTGGCGTTCACCGAGCCCGCAGCGTGTGGCTGACACTCTGGGCGCGTCTGTTACGCCGAG	1200
Db	1141	GTGGCGTTCACCGAGCCCGCAGCGTGTGGCTGACACTCTGGGCGCGTCTGTTACGCCGAG	1200
QY	1201	CACAGTCGGGAGTAAAGGAAAGCATGA	1228
Db	1201	CACAGTCGGGAGTAAAGGAAAGCATGA	1228
RESULT	8		
LOCUS	A87606	1235 bp	DNA
DEFINITION	Sequence 3 from Patent WO936089.		PAT
ACCESSION	A87606		22-JAN-2000
VERSION	A87606.1	GI:6736246	
KEYWORDS			
SOURCE	unidentified.		
ORGANISM	unidentified.		
REFERENCE	1 (bases 1 to 1235)		
AUTHORS	Flohe, L. and Singh, M.		
TITLE	TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE		
JOURNAL	Patent: WO 9836089-A-20-AUG-1998; FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)		
FEATURES	Location/Qualifiers		
Source	1..1235		
BASE COUNT	236 a 394 c 382 g 220 t 3 others		
ORIGIN	/db_xref="taxon:32644"		
Query Match	99.7%, Score 1224.4; DB 5; Length 1235;		
Best Local Similarity	99.8%; Pred. No. 1.4e-169;		
Matches 1225; Conservative	0; Mismatches 3; Indels 0; Gaps 0;		
QY	1	ATCTTGCAAGATTATCGAAGCTTCTTCATCTACGAGCGTACGATTCGAGAGGGTAATC	60
Db	1	ATCTTGCAAGATTATCGAAGCTTCTTCACACTGAGCGTACGATTCGAGAGGGTAATC	60
QY	61	ATGCGCGTGTGATTTCGACGACGCAAAAACAAGAAATTCGGGTGGCATCAACCCG	120
Db	61	ATGCGCGTGTGATTTCGACGACGCAAAAACAAGAAATTCGGGTGGCATCAACCCG	120
QY	121	GGCGGCGTGGCGGAAGTAACCCGCTGTGGCCATGAGTCTCATCCAGGACAGTGGCGGA	180
Db	121	GGCGGCGTGGCGGAAGTAACCCGCTGTGGCCATGAGTCTCATCCAGGACAGTGGCGGA	180
QY	181	GAGGCGTGGCTATACACGACCGGATTTCAAGAGGGGCGAGCGCGAATGGTTCGGCACC	240
Db	181	GAGGCGTGGCTATACACGACCGGATTTCAAGAGGGGCGAGCGCGAATGGTTCGGCACC	240
QY	241	GCGGACCAAGTGTGGGCGGACGCTGATTTATTGCTCAAGGTCAAAAGAACGATAGCGGCG	300
Db	241	GCGGACCAAGTGTGGGCGGACGCTGATTTATTGCTCAAGGTCAAAAGAACGATAGCGGCG	300
QY	301	GATATGCGGCGGCTGGGAAGGGGAGATCTTGTTACGTTCTTGATTTGGCGGCGTCA	360
Db	301	GATATGCGGCGGCTGGGAAGGGGAGATCTTGTTACGTTCTTGATTTGGCGGCGTCA	360
QY	361	CGTGTTCGACCGATCGCTGTGTGATTCGGGACACCACTCAATTCCTACGAGACCGTC	420
Db	361	CGTGTTCGACCGATCGCTGTGTGATTCGGGACACCACTCAATTCCTACGAGACCGTC	420
QY	421	CAGACCGCGGAGGGGCACTACCCCTGCTTGGCCCGATAGAGAAATGCCCGGTGCATTC	480

```

      Db 421 CAGACCGCCGACGCGGCTACCTCCCTGCTTGCCTCCGATGACGAACTGCGCGGTGACTC 480
      Oy 481 GCGCCGACGTTGGCGCTTACCACTGATGCGAACCAGAGGGGCGCGGTGTGCTGATG 540
      Db 481 GCGCCGACGTTGGCGCTTACCACTGATGCGAACCAGAGGGGCGCGGTGTGCTGATG 540
      Oy 541 GCGCGGCTGCGCGCTGCGAAGCGCGGCGGAGTGTGTATGCGCGCGGACCGCGGCG 600
      Db 541 GCGCGGCTGCGCGCTGCGAAGCGCGGCGGAGTGTGTATGCGCGCGGACCGCGGCG 600
      Oy 601 TACAAAGCAGCGCGGATGCGCAAGCGGATGCGGCGGATGCGGCTTACGATCAAC 660
      Db 601 TACAAAGCAGCGCGGATGCGCAAGCGGATGCGGCGGATGCGGCTTACGATCAAC 660
      Oy 661 ATGACAAATTTGGGCACTGCAAGCGGATGCGGCGGATGCGGCTTACGATCAAC 720
      Db 661 ATGACAAATTTGGGCACTGCAAGCGGATGCGGCGGATGCGGCTTACGATCAAC 720
      Oy 721 TCGGCTACGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGATGCGGCTTACGATCAAC 780
      Db 721 TCGGCTACGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGATGCGGCTTACGATCAAC 780
      Oy 781 GTGCGAGCGCGCAAGCGGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATCAAC 840
      Db 781 GTGCGAGCGCGCAAGCGGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATCAAC 840
      Oy 841 GTGCGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGATGCGGCTTACGATGCGGCTTACGATCAAC 900
      Db 841 GTGCGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGATGCGGCTTACGATGCGGCTTACGATCAAC 900
      Oy 901 ACCACCTACGAGCGGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATCAAC 960
      Db 901 ACCACCTACGAGCGGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATCAAC 960
      Oy 961 ATGCGCGCTGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATCAAC 1020
      Db 961 ATGCGCGCTGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATCAAC 1020
      Oy 1021 GTGCGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGATGCGGCTTACGATGCGGCTTACGATCAAC 1080
      Db 1021 GTGCGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGATGCGGCTTACGATGCGGCTTACGATCAAC 1080
      Oy 1081 AAAGGTCTTGGAGCGAGGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATCAAC 1140
      Db 1081 AAAGGTCTTGGAGCGAGGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATCAAC 1140
      Oy 1141 GTGCGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGATGCGGCTTACGATGCGGCTTACGATCAAC 1200
      Db 1141 GTGCGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGATGCGGCTTACGATGCGGCTTACGATCAAC 1200
      Oy 1201 CACACGTGCGAGGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATCAAC 1260
      Db 1201 CACACGTGCGAGGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATCAAC 1260
  
```

```

RESULT 9
LOCUS A89746 1235 bp DNA
DEFINITION Sequence 3 from Patent WO932862.
ACCESSION A89746
VERSION A89746.1 GI:6738280
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1235)
AUTHORS Flohe, L. and Singh, M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 932862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
source 1..1235
/organism="unidentified"
  
```

```

BASE COUNT 236 a 394 c 382 g 220 t 3 others
ORIGIN
Query Match 99.7%; Score 1224.4; DB 5; Length 1235;
Best Local Similarity 99.8%; Pred. No. 1.4e-168;
Matches 1225; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
      Oy 1 ATCTTGCAGATTATTCAGACTTTCTTCACTACTGAAAGCTTACAGTATGAGAGGGGATATC 60
      Db 1 ATCTTGCAGATTATTCAGACTTTCTTCACTACTGAAAGCTTACAGTATGAGAGGGGATATC 60
      Oy 61 ATGCGCGTGGGATTCGACCGAGAGCAAAAACAGATTCGCGGTGGCCATACCCCG 120
      Db 61 ATGCGCGTGGGATTCGACCGAGAGCAAAAACAGATTCGCGGTGGCCATACCCCG 120
      Oy 121 GCGGCGTGGGGAATTAACCGGTGTCGATGAGGTGCTATTCAGAGCAAGTGGCGGA 180
      Db 121 GCGGCGTGGGGAATTAACCGGTGTCGATGAGGTGCTATTCAGAGCAAGTGGCGGA 180
      Oy 181 GAGGCGTGGGCTATACCGAGCGGATTCGAAAGCGGCGGCGCACTGTGGCGAGC 240
      Db 181 GAGGCGTGGGCTATACCGAGCGGATTCGAAAGCGGCGGCGCACTGTGGCGAGC 240
      Oy 241 GCGGACGAGGTGGGCGGAGCGGCTGATTTATGCTCAAGGTCAAGAACCGATAGCGCG 300
      Db 241 GCGGACGAGGTGGGCGGAGCGGCTGATTTATGCTCAAGGTCAAGAACCGATAGCGCG 300
      Oy 301 GAATACGGCGGCTGCGACAGCGGCGATCTTGTTCACGTTCTTGCATTTGGCGGCTCA 360
      Db 301 GAATACGGCGGCTGCGACAGCGGCGATCTTGTTCACGTTCTTGCATTTGGCGGCTCA 360
      Oy 361 CGTGTGACAGGATGCGTGTGTTGATTCGCGGACGACGATGCTCAATGGCTACGAGACGCTC 420
      Db 361 CGTGTGACAGGATGCGTGTGTTGATTCGCGGACGACGATGCTCAATGGCTACGAGACGCTC 420
      Oy 421 CAGACCGCGGAGCGGCGACATACCGCTGCTTGGCGCGATGAGCGAAGTGGCGGTGACTC 480
      Db 421 CAGACCGCGGAGCGGCGACATACCGCTGCTTGGCGCGATGAGCGAAGTGGCGGTGACTC 480
      Oy 481 GCGGCGGAGTGGGCTTACCACTGATGCGAACCAGAGGGGCGGCGGTGTGATG 540
      Db 481 GCGGCGGAGTGGGCTTACCACTGATGCGAACCAGAGGGGCGGCGGTGTGATG 540
      Oy 541 GCGGCGGAGTGGGCTTACCACTGATGCGAACCAGAGGGGCGGCGGTGTGATG 600
      Db 541 GCGGCGGAGTGGGCTTACCACTGATGCGAACCAGAGGGGCGGCGGTGTGATG 600
      Oy 601 TACACGAGCGCGCATGCGCAAGCGGATGCGGCGGATGCGGCTTACGATGCGGCTTACGATCAAC 660
      Db 601 TACACGAGCGCGCATGCGCAAGCGGATGCGGCGGATGCGGCTTACGATGCGGCTTACGATCAAC 660
      Oy 661 ATGACAAATTTGGGCACTGCAAGCGGATGCGGCGGATGCGGCTTACGATGCGGCTTACGATCAAC 720
      Db 661 ATGACAAATTTGGGCACTGCAAGCGGATGCGGCGGATGCGGCTTACGATGCGGCTTACGATCAAC 720
      Oy 721 TCGGCTACGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGATGCGGCTTACGATGCGGCTTACGATCAAC 780
      Db 721 TCGGCTACGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGATGCGGCTTACGATGCGGCTTACGATCAAC 780
      Oy 781 GTGCGAGCGCGCAAGCGGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATCAAC 840
      Db 781 GTGCGAGCGCGCAAGCGGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATCAAC 840
      Oy 841 GTGCGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGATGCGGCTTACGATGCGGCTTACGATCAAC 900
      Db 841 GTGCGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGATGCGGCTTACGATGCGGCTTACGATCAAC 900
      Oy 901 ACCACCTACGAGCGGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATCAAC 960
      Db 901 ACCACCTACGAGCGGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATGCGGCTTACGATCAAC 960
  
```

```

OY 961 ATGCCCCGCTGGTGCCGAGAGAGCTGACCTACGGGCTGACCAACGAGCATGCCGTAT 1020
DB 961 ATGCCCCGCTGGTGCCGAGAGAGCTGACCTACGGGCTGACCAACGAGCATGCCGTAT 1020
OY 1021 GTGCTGAGACTTGCAGACCATGGCTGGCGGGGCTGGCGGTGCAATCCGGCTAGCC 1080
DB 1021 GTGCTGAGACTTGCAGACCATGGCTGGCGGGGCTGGCGGTGCAATCCGGCTAGCC 1080
OY 1081 AAGAGCTTTGACAGCAGCAAGAGGGCTTACCTGCGAAGGGGTGGCCAGCACTGGGG 1140
DB 1081 AAGAGCTTTGACAGCAGCAAGAGGGCTTACCTGCGAAGGGGTGGCCAGCACTGGGG 1140
OY 1141 GTGCGCTTACAGAGAGCCGCGAGGCTGTGGCTGACTCTCGGCGCTGTTACCGGAG 1200
DB 1141 GTGCGCTTACAGAGAGCCGCGAGGCTGTGGCTGACTCTCGGCGCTGTTACCGGAG 1200
OY 1201 CACAGCTCGGAGTAAGGAGACGATGA 1228
DB 1201 CACAGCTCGGAGTAAGGAGACGATGA 1228

RESULT 10
LOCUS A87609 1235 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 6 from Patent WO9836089.
ACCESSION A87609
VERSION A87609.1 GI:6736249
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1235)
AUTHORS Flohe, U. and Singh, M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-Aug-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
Source Location/Qualifiers
1..1235
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 235 a 395 c 384 g 220 t 1 others
ORIGIN
Query Match 98.8%; Score 1213.4; DB 5; Length 1235;
Best Local Similarity 99.8%; Pred. No. 5.4e-167;
Matches 1225; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```

```

DB 360 CGTGCTTGACAGATGGGTGTGATTCGGGCAACAGTCAATTGGCTAGAGACCGTC 419
OY 421 CAGACCGCGCAGCGGACACTACCCCTGCTTGCCTCCGATGACGAAATGCGCGTCACTC 480
DB 420 CAGACCGCGCAGCGGACACTACCCCTGCTTGCCTCCGATGACGAAATGCGCGTCACTC 479
OY 481 GCGCGCCAGGTGGGCTTACCACTGATGCGAAGCCCAAGGGGACCGGGTGTCTATG 540
DB 480 GCGCGCCAGGTGGGCTTACCACTGATGCGAAGCCCAAGGGGACCGGGTGTCTATG 539
OY 541 GCGGGGTGCCCCGCTGCAACCGGCGAGCTGCTGTGATCGAGCGCCGCGCCGCG 600
DB 540 GCGGGGTGCCCCGCTGCAACCGGCGAGCTGCTGTGATCGAGCGCCGCGCCGCG 599
OY 601 TACAAGCAGCGCGCATGCGCAAGGGCATGGGGGAGACGTTAGGGTTCTAGACATCAAC 660
DB 600 TACAAGCAGCGCGCATGCGCAAGGGCATGGGGGAGACGTTAGGGTTCTAGACATCAAC 659
OY 661 ATGCAAAACTTTCGCAACTGAGACCGCGAGTTCTGCGGCGGATTCACACTGCTACTCA 720
DB 660 ATGCAAAACTTTCGCAACTGAGACCGCGAGTTCTGCGGCGGATTCACACTGCTACTCA 719
OY 721 TCGGCTTACAGACTGAGAGGCTGCCGTCAAAAGCTGCGCACTGGGTATGGGGCGTCTG 780
DB 720 TCGGCTTACAGACTGAGAGGCTGCCGTCAAAAGCTGCGCACTGGGTATGGGGCGTCTG 779
OY 781 GTGCAGAGCGCAAGGCAACCAATTAATCTGCAATTACTTGTGCGCATATGAACA 840
DB 780 GTGCAGAGCGCGCAAGGCAACCAATTAATCTGCAATTACTTGTGCGCATATGAACA 839
OY 841 GGTGCGGTACTGTGTGATATAGCATTCAGACCAAGGGCGCTGTTGAAAGCTCACAGACG 900
DB 840 GGTGCGGTACTGTGTGATATAGCATTCAGACCAAGGGCGCTGTTGAAAGCTCACAGACG 899
OY 901 ACCACTTACAGACCAACCGGAGCTGCGGCGAGCAGCAGCTGTTTACCTGGGCGAGAC 960
DB 900 ACCACTTACAGACCAACCGGAGCTGCGGCGAGCAGCAGCTGTTTACCTGGGCGAGAC 959
OY 961 ATGCCCCGCTGGTGCCGAGAGAGCTGACCTAGCGGCTGACCAACGAGCATGCCGTAT 1020
DB 960 ATGCCCCGCTGGTGCCGAGAGAGCTGACCTAGCGGCTGACCAACGAGCATGCCGTAT 1019
OY 1021 GTGCTGAGACTTGCAGACCATGGCTGGCGGGGCTGGCGGTGCAATCCGGCTAGCC 1080
DB 1020 GTGCTGAGACTTGCAGACCATGGCTGGCGGGGCTGGCGGTGCAATCCGGCTAGCC 1079
OY 1081 AAGAGCTTTGACAGCAGCAAGAGGGCTTACTGTGCGAAGGGGTGGCCAGCACTGGGG 1140
DB 1080 AAGAGCTTTGACAGCAGCAAGAGGGGCTTACTGTGCGAAGGGGTGGCCAGCACTGGGG 1139
OY 1141 GTGCGCTTACAGAGAGCCGCGAGGCTGTGGCTTACCTGCGGCTGTGAGCCGAG 1200
DB 1140 GTGCGCTTACAGAGAGCCGCGAGGCTGTGGCTTACCTGCGGCTGTGAGCCGAG 1199
OY 1201 CACAGCTCGGAGTAAGGAGACGATGA 1228
DB 1200 CACAGCTCGGAGTAAGGAGACGATGA 1227

RESULT 11
LOCUS A89749 1235 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 6 from Patent WO9832862.
ACCESSION A89749
VERSION A89749.1 GI:6738283
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1235)
AUTHORS Flohe, U. and Singh, M.
TITLE L-7-ANINE DEHYDROGENASE OF MYCOBACTERIUM MARINDU

```



**JOURNAL**  
**Patent: WO 9832862-A 30-JUL-1998;**  
**FIORIE TEODORO (DE) · SINGH MAHAVIR**

FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE);  
Tschertanov/Qualifiers

source	1. .1235
--------	----------

	source
A.I.: 1289 /organism	

```

/organism="unidentified"
/db_xref="taxon:32644"

```

ASE COUNT	235 a	395 c	384 g	220 t	1 others

RIGIN

ORIGIN

Query Match	98.8%	Score 1213.4;	DB 5;	Length 1235;
Best Local Similarity	99.8%	Pred. No. 5.4e-167;		
Matches 1225; Conservative	0;	Mismatches 2;	Indels 1;	Gaps 1

QY	1	ATCTTGAGATTATATCGAACTTCTTCATCTAGTAAAGGTACACATATCGAGAGGGGTATTC	60
Db	1	ATCTTGAGATTATATCGAACTTCTTCATCTAGTAAAGGTACACATATCGAGAGGGGTATTC	60
QY	61	ATGCGCGTCCGGTATTCGCCAGCCGAGACCAAAACAACT	120
Db	61	ATGCGCGTCCGGTATTCGCCAGCCGAGACCAAAACAACT	120
QY	121	GCGGCGTCCGGGAACTTAACCCGTCGTGGGCGATGAGGTGCTCATCGACGAGGTGCGGGA	180
Db	121	GCGGCGTCCGGGAACTTAACCCGTCGTGGGCGATGAGGTGCTCATCGACGAGGTGCGGGA	180
QY	181	GAGGCGTCCGGTATTCACCGAGCGCGGATTTCAAGGCGGCGCGCACTG	240
Db	181	GAGGCGTCCGGTATTCACCGAGCGCGGATTTCAAGGCGGCGCGCACTG	240
QY	241	GCCGACCAAGTGTGGGCGCGAGCCGTGATTTATGCTCAAGGTCAAAGAACCGATAGGGGGG	300
Db	241	GCCGACCAAGTGTGGGCGCGAGCCGTGATTTATGCTCAAGGTCAAAGAACCGATAGGGGGG	300
QY	301	GAATACGGCCCTCTGCGACACAGGGCAATTTGTTCACTTCTTGATTTGGCGCGGTCA	360
Db	301	GAATACGGCCCTCTGCGACACAGGGCAATTTGTTCACTTCTTGATTTGGCGCGGTCA	360
QY	361	CGTCTTTCACCCGATTCGGTGTGTGGATTCGGGACACACGTCATTTGCTTACGAGACCGTC	420
Db	361	CGTCTTTCACCCGATTCGGTGTGTGGATTCGGGACACACGTCATTTGCTTACGAGACCGTC	420
QY	421	CAGACCGCGCGAGCGGCGGCACTACCCCTCTTGCCCCGATGAGCAAGTCGCGGTCGACTC	480
Db	421	CAGACCGCGCGAGCGGCGGCACTACCCCTCTTGCCCCGATGAGCAAGTCGCGGTCGACTC	480
QY	481	GCCGCCAGTGTGGCGCTTACCACTGATGCGAACCACCAAGGGGCGCGGTGTGCTGATG	540
Db	481	GCCGCCAGTGTGGCGCTTACCACTGATGCGAACCACCAAGGGGCGCGGTGTGCTGATG	540
QY	541	GCGGCGGCGCGCGCGCTTCGAAACGCGCGAGCGTGTGATGATCGGCGCGCGGACCGCGCGC	600
Db	541	GCGGCGGCGCGCGCGCTTCGAAACGCGCGAGCGTGTGATGATCGGCGCGCGGACCGCGCGC	600
QY	601	TACAAACGACGCGCGATTCGCGCAAGCGCATGGGCGCGACCGTTACGCTTACAGATCAAC	660
Db	601	TACAAACGACGCGCGATTCGCGCAAGCGCATGGGCGCGACCGTTACGCTTACAGATCAAC	660
QY	661	ATCGACAAATCTTCGCAACTGACGCGGAGTTCGTGGGCGCGGATTCACACTCGCTATCTCA	720
Db	661	ATCGACAAATCTTCGCAACTGACGCGGAGTTCGTGGGCGCGGATTCACACTCGCTATCTCA	720
QY	721	TGCGGCTACGAGCTGAGAGGTGCGGTCAAAAGTGCAGACTGTGATTTGGGCGCGTCTCTG	780
Db	721	TGCGGCTACGAGCTGAGAGGTGCGGTCAAAAGTGCAGACTGTGATTTGGGCGCGTCTCTG	780
QY	781	GTGCCAGGCGCCAGGACCCCAATTAAGTCTGCAATTCCTGTTCGCGCATATTAACA	840
Db	781	GTGCCAGGCGCCAGGACCCCAATTAAGTCTGCAATTCCTGTTCGCGCATATTAACA	840
QY	841	GGTGGGCTACGTGTGATATATGCAATTCACACAGGCGCGCTTTTGAAAGGCTCAGACCG	900
Db	841	GGTGGGCTACGTGTGATATATGCAATTCACACAGGCGCGCTTTTGAAAGGCTCAGACCG	900

QY	901	ACACACATACACACACCCGAGCTTCCGCTGACAGACAGCTTTTACTGCTGGCGAAC	960
Db	900	ACCACTACAGACACACCCGAGCTTCCGCTGACAGACAGCTTTTACTGCTGGCGAAC	959
QY	961	ATGCGCCGCTCGTGGCCGAAGACGTGACCTACGGGCTGACCAAGGAGCATGCCGAT	1020
Db	960	ATGCGCCGCTCGTGGCCGAAGACGTGACCTACGGGCTGACCAAGGAGCATGCGCAT	1019
QY	1021	GTCCTCAGATTGCCGCATATGCTGGCGGGCGCGTGGCGCTCGAATCGGACACTAACCC	1080
Db	1020	GTCCTCAGAGTTGCCGCACATATGCTGGCGGGCGCGTGGCGCTCGAATCGGACACTAACCC	1079
QY	1081	AAAGCTTTTTCGACGACGACGAAAGGGGGCTTACTGTCCGAAGGGGTGGCCACCGACTGGGG	1140
Db	1080	AAAGCTTTTTCGACGACGACGAAAGGGGGCTTACTGTCCGAAGGGGTGGCCACCGACTGGGG	1139
QY	1141	GTCGCCGTTCACCGAGCCCGCCAGCGTGTGGCCCTGACTCTCGGCGCTGTGTTACGCGAG	1200
Db	1140	GTCGCCGTTCACCGAGCCCGCCAGCGTGTGGCCCTGACTCTCGGCGCTGTGTTACCGCGAAN	1199
QY	1201	CACACGTGGGAGTAAGGAAAGCGATGA 1228	
Db	1200	CACACGTGGGAGTAAGGAAAGCGATGA 1227	

[illegible]

```

QY 301 GAATACGGCGCGCTGCGACGAGGAGATCTTGTACAGTCTTGATTTGGCCGGCTCA 360
    |||||||
Db 301 GAATACGGCGCGCTGCGACGAGGCG -GATCTTGTACAGTCTTGATTTGGCCGGCTCA 359
QY 361 CGTGTGACCGCATCGCTTTGTGATTCGGCAGACAGTCAATTTCCCTACAGACCGTC 420
    |||||||
Db 361 CGTGTGACCGCATCGCTTTGTGATTCGGCAGACAGTCAATTTCCCTACAGACCGTC 419
QY 421 CAAACCGCGAGGCGGACATACCCCTGCTTGGCCGATGAGCAAGTCCGGTGCATC 480
    |||||||
Db 420 CAAACCGCGAGGCGGACATACCCCTGCTTGGCCGATGAGCAAGTCCGGTGCATC 479
QY 481 GCGCGCCAGGTTGGCGCTTACACACTGATGCGAACCCAGGGGGCGCGGTGTCTGATG 540
    |||||||
Db 480 GCGCGCCAGGTTGGCGCTTACACACTGATGCGAACCCAGGGGGCGCGGTGTCTGATG 539
QY 541 GGGGGGTGCGCGCGCGTGAACCGGCGAGCGTGTGTGATGCGCGCGGACCGCGCGG 600
    |||||||
Db 540 GGGGGGTGCGCGCGCGTGAACCGGCGAGCGTGTGTGATGCGCGCGGACCGCGCGG 599
QY 601 TACAACGACCGCGCATCGCCAAAGCGCATGGCGCGACCTTACAGTTCAGACATCAAC 660
    |||||||
Db 600 TACAACGACCGCGCATCGCCAAAGCGCATGGCGCGACCTTACAGTTCAGACATCAAC 659
QY 661 ATGACAAACTTGGGCACTGCAAGCGGAGTTCGCGCGCGGATCCACACTGCTACTCA 720
    |||||||
Db 660 ATGACAAACTTGGGCACTGCAAGCGGAGTTCGCGCGCGGATCCACACTGCTACTCA 719
QY 721 TCGGCTACGAGCTCGAGGGTGGCGTAAACGTGCGGAGTGTGATTTGGCGCGTCTG 780
    |||||||
Db 720 TCGGCTACGAGCTCGAGGGTGGCGTAAACGTGCGGAGTGTGATTTGGCGCGTCTG 779
QY 781 GTGCCAGCGCGCAAGGACACCAATTTAGTCTGCAATCACTTGGCGCGCATATGAACG 840
    |||||||
Db 780 GTGCCAGCGCGCAAGGACACCAATTTAGTCTGCAATCACTTGGCGCGCATATGAACG 839
QY 841 GGTGCGTACTGTGATATAGCCATGACGACGAGGCGGTGTTTCGAAGGCTCAGACCG 900
    |||||||
Db 840 GGTGCGTACTGTGATATAGCCATGACGACGAGGCGGTGTTTCGAAGGCTCAGACCG 899
QY 901 ACCACCTACGACACCGGAGCTTGGCGTGCAGACACAGCTGTTTACTCGTGGCGAAC 960
    |||||||
Db 900 ACCACCTACGACACCGGAGCTTGGCGTGCAGACACAGCTGTTTACTCGTGGCGAAC 959
QY 961 ATGGCGCGCTGTGTCGCGAAGCTGACGCTACGCGCTGACCAAGCGAGATCCGCTAT 1020
    |||||||
Db 960 ATGGCGCGCTGTGTCGCGAAGCTGACGCTACGCGCTGACCAAGCGAGATCCGCTAT 1019
QY 1021 GTGCTCGAGCTTGGCGACCATGCTGGCGGGCGCGTGGTGAATCCGGCACTAGCC 1080
    |||||||
Db 1020 GTGCTCGAGCTTGGCGACCATGCTGGCGGGCGCGTGGTGAATCCGGCACTAGCC 1079
QY 1081 AAAGTCTTTCGAGGACGAGAGGGGCGTTACTGTCCGAAGGGTGGGACGACGACTGGG 1140
    |||||||
Db 1080 AAAGTCTTTCGAGGACGAGAGGGGCGTTACTGTCCGAAGGGTGGGACGACGACTGGG 1139
QY 1141 GTGCGCTTACCGAGCCCGCGGAGCGTGTGCTGACTCTGGCGCGCTCGTTACGCGGAG 1200
    |||||||
Db 1140 GTGCGCTTACCGAGCCCGCGGAGCGTGTGCTGACTCTGGCGCGCTCGTTACGCGGAG 1199
QY 1201 CACACGTGGGAGTAAAGGAAGCATGA 1228
    |||||||
Db 1200 CACACGTGGGAGTAAAGGAAGCATGA 1227

```

## RESULT 13

```

LOCUS      A89751      1235 bp      DNA
DEFINITION Sequence 8 from Patent W09832862.
ACCESSION A89751
VERSION   A89751.1   GI:6738285
KEYWORDS
SOURCE      unidentified.

```

```

ORGANISM      unidentified
REFERENCE      1 (bases 1 to 1235)
AUTHORS      Flohe, L. and Singh, M.
TITLE      L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL      Patent: WO 9832862-A 30-JUL-1998;
              FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
  source
    1..1235
    /organism="unidentified"
    /db_xref="taxon:32644"
  BASE COUNT      236 a      394 c      385 g      220 t
  ORIGIN
Query Match      98.8%; Score 1212.8; DB 5; Length 1235;
Best Local Similarity 99.8%; Pred. No. 6.6e-167;
Matches 1225; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 ATCTGCAAGATTATGCAACTTCTCTACTGAAAGCGTACGATGAGAGGGGTAATC 60
    |||||||
Db 1 ATCTGCAAGATTATGCAACTTCTCTACTGAAAGCGTACGATGAGAGGGGTAATC 60
QY 61 ATGCGCGTGGTATTCGACGAGACCAAAAAACAAGAAATTCGGGTGGCCATACCCCG 120
    |||||||
Db 61 ATGCGCGTGGTATTCGACGAGACCAAAAAACAAGAAATTCGGGTGGCCATACCCCG 120
QY 121 GCGGCGTCGCGGAACCTAACCCGTCGTGGCCATGAGTGTGCTATCCAGGCGAGTGGCGGA 180
    |||||||
Db 121 GCGGCGTCGCGGAACCTAACCCGTCGTGGCCATGAGTGTGCTATCCAGGCGAGTGGCGGA 180
QY 181 GAGGGCTGGCTATACCGGACGCGGATTTCAAGGGCGGCGGCAACTGGTGGCAC 240
    |||||||
Db 181 GAGGGCTGGCTATACCGGACGCGGATTTCAAGGGCGGCGGCAACTGGTGGCAC 240
QY 241 GCGGACGAGGTGGGCGGACGCTATTTTGTCTCAAGGTCAAGAAACGATAGCGCG 300
    |||||||
Db 241 GCGGACGAGGTGGGCGGACGCTATTTTGTCTCAAGGTCAAGAAACGATAGCGCG 300
QY 301 GAATACGGCGCGCTGCGACGAGGCGAGATCTTGTTCAGCTTCTGATTTGGCGCGCTCA 360
    |||||||
Db 301 GAATACGGCGCGCTGCGACGAGGCGAGATCTTGTTCAGCTTCTGATTTGGCGCGCTCA 359
QY 361 CGTGTGACCGCATCGCTTTGTGATTCGGCAGACAGTCAATTTCCCTACAGACCGTC 420
    |||||||
Db 361 CGTGTGACCGCATCGCTTTGTGATTCGGCAGACAGTCAATTTCCCTACAGACCGTC 419
QY 421 CAAACCGCGAGGCGGACATACCCCTGCTTGGCCGATGAGCAAGTCCGGTGCATC 480
    |||||||
Db 420 CAAACCGCGAGGCGGACATACCCCTGCTTGGCCGATGAGCAAGTCCGGTGCATC 479
QY 481 GCGCGCCAGGTTGGCGCTTACACACTGATGCGAACCCAGGGGGCGGCTGTGATG 540
    |||||||
Db 480 GCGCGCCAGGTTGGCGCTTACACACTGATGCGAACCCAGGGGGCGGCTGTGATG 539
QY 541 GCGGGGTGCGCGCGGCTGCAACCGGCGGAGCTGTGTATGGGGCGGCGACCGCGCG 600
    |||||||
Db 540 GCGGGGTGCGCGCGGCTGCAACCGGCGGAGCTGTGTATGGGGCGGCGACCGCGCG 599
QY 601 TACAACGACCGCGCATCGCCAAAGCGCATGGGCGGACCGTACGTTCTAGACATCAAC 660
    |||||||
Db 600 TACAACGACCGCGCATCGCCAAAGCGCATGGGCGGACCGTACGTTCTAGACATCAAC 659
QY 661 ATGACAAACTTGGGCACTGCAAGCGGAGTTCGCGCGCGGATCCACACTGCTACTCA 720
    |||||||
Db 660 ATGACAAACTTGGGCACTGCAAGCGGAGTTCGCGCGCGGATCCACACTGCTACTCA 719
QY 721 TCGGCTACGAGCTCGAGGGTGGCGTCAACGTCGCGGACGCTGTATTTGGCGCGTCTG 780
    |||||||
Db 720 TCGGCTACGAGCTCGAGGGTGGCGTCAACGTCGCGGACGCTGTATTTGGCGCGTCTG 779
QY 781 GTGCCAGCGCGCAAGGACACCAATTTAGTCTGCAATCACTTGGCGCGCATATGAACCA 840
    |||||||

```

Db 780 GTGCCAGCGCCGCAAGGCAACCAATAGTCTCGAATTCACATTGTGCGCATATGAACCA 839  
 QY 841 GGTGCGTACTGTGATATATACCATGACAGGCGGCTGTTTCGAAGGCTCAGCACCG 900  
 Db 840 GGTGCGTACTGTGATATATAGCATGACAGGCGGCTGTTTCGAAGGCTCAGCACCG 899  
 QY 901 ACACGCTACGACGACGCGGAGCTTGGCGTGACAGACGCGTGTCTTACTGCTGGCGAAC 960  
 Db 900 ACACGCTACGACGACGCGGAGCTTGGCGTGACAGACGCGTGTCTTACTGCTGGCGAAC 959  
 QY 961 ATGCGCGCTGCGTGGCGAAGACGTCGACCTACGCGCTGACCAACGCGAGATGCGTAT 1020  
 Db 960 ATGCGCGCTGCGTGGCGAAGACGTCGACCTACGCGCTGACCAACGCGAGATGCGTAT 1019  
 QY 1021 GTGCTGAGCTTGGCGACGATAGGCTGGCGGCGGCGGCGGCTGGATCCGCGACTAGCC 1080  
 Db 1020 GTGCTGAGCTTGGCGACGATAGGCTGGCGGCGGCGGCGGCTGGATCCGCGACTAGCC 1079  
 QY 1081 AAAGGTCTTTCGACGACGAGAGGCGCTTACTGTCCGAACGGGTGGCCACCGACCTGGG 1140  
 Db 1080 AAAGGTCTTTCGACGACGAGAGGCGCTTACTGTCCGAACGGGTGGCCACCGACCTGGG 1139  
 QY 1141 GTGCGCTTACCGACGCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
 Db 1140 GTGCGCTTACCGACGCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1199  
 QY 1201 CACAGCTGGGAGTAAGGAGGAGCATGA 1228  
 Db 1200 CACAGCTGGGAGTAAGGAGGAGCATGA 1227

## RESULT 14

LOCUS AB87610 1229 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 7 from Patent WO9836089.  
 ACCESSION AB87610  
 VERSION AB87610.1 GI:6736250  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified  
 unclassified.  
 REFERENCE  
 1. (bases 1 to 1229)  
 Flohe, L. and Singh, M.  
 TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 Patent: WO 9836089-A 20-AUG-1998;  
 JOURNAL FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 Location/Qualifiers  
 1..1229  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 236 a 391 c 382 g 219 t 1 others  
 ORIGIN

Query Match 98.7%: Score 1211.8; DB: 5; Length 1229;  
 Best Local Similarity 99.7%: Pred. No. 9.3e-167;

Matches 1224: Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 ATCTGCGAGATTATGCAATCTTCTTCACTGAGAGCGTATCGAGAGGGGTATC 60  
 Db 1 ATCTGCGAGATTATGCAATCTTCTTCACTGAGAGCGTATCGAGAGGGGTATC 60  
 QY 61 ATGCGGCTGGATTTCGACGACGACCAAAACAGATTCCGGGTGCGCATCACCCG 120  
 Db 61 ATGCGGCTGGATTTCGACGACGACCAAAACAGATTCCGGGTGCGCATCACCCG 120  
 QY 121 GCGCGGCTGGCGAATAACCGCTGCGCATGAGTCTCTATCCAGCAGGTGCGGGA 180  
 Db 121 GCGCGGCTGGCGAATAACCGCTGCGCATGAGTCTCTATCCAGCAGGTGCGGGA 180  
 QY 181 GAGGCTGCGCTATCACGACGCGGATTTCAAGGCGGCGCGCAACTGGTGGGACC 240  
 Db 181 GAGGCTGCGCTATCACGACGCGGATTTCAAGGCGGCGCGCAACTGGTGGGACC 240

QY 241 GCGGACGAGGTGAGGCGGCGAGCTGATTTATTCGCAAGGTCAAGAACCGATAGCGCG 300  
 Db 241 GCGGACGAGGTGAGGCGGCGAGCTGATTTATTCGCAAGGTCAAGAACCGATAGCGCG 300  
 QY 301 GAATACGCGCGCTGCGACACGCGGATATCTTGTACGCTTCTTGCATTTGGCGCGTCA 360  
 Db 301 GAATACGCGCGCTGCGACACGCGGATATCTTGTACGCTTCTTGCATTTGGCGCGTCA 359  
 QY 361 CGTCTTTCACGAGTACGCTTGTGATTCGCGACACGCTCAATTTGCTTACGACGCTC 420  
 Db 360 CGTCTTTCACGAGTACGCTTGTGATTCGCGACACGCTCAATTTGCTTACGACGCTC 419  
 QY 421 CAGACCGCGGACGCGCATATACCCGCTTGGCGCGATGAGGAAAGTGGCGGCTGACTC 480  
 Db 420 CAGACCGCGGACGCGCATATACCCGCTTGGCGCGATGAGGAAAGTGGCGGCTGACTC 479  
 QY 481 GCGCGCGAGTTGGCGCTTACGACCTGATGCGAACCCAAAGGCGCGGCTGTGCTATG 540  
 Db 480 GCGCGCGAGTTGGCGCTTACGACCTGATGCGAACCCAAAGGCGCGGCTGTGCTATG 539  
 QY 541 GCGCGGCTGCGCGGCTGCGAACCGCGCGAGCTGTGATGCGGCGCGGCGGCGGCG 600  
 Db 540 GCGCGGCTGCGCGGCTGCGAACCGCGCGAGCTGTGATGCGGCGCGGCGGCGGCG 599  
 QY 601 TACAGCGACGCGCGCATGCGCAAGGCGCATGCGGCGGCGGCTTACGCTTACGATCAAC 660  
 Db 600 TACAGCGACGCGCGCATGCGCAAGGCGCATGCGGCGGCGGCTTACGCTTACGATCAAC 659  
 QY 661 ATGACAACTTGGCGCATCTGACGCGCGAGCTTGGCGCGGATCCACCTGCTACTCA 720  
 Db 660 ATGACAACTTGGCGCATCTGACGCGCGAGCTTGGCGCGGATCCACCTGCTACTCA 719  
 QY 721 TCGGCTCAGAGCTCGAGGCTGCGTCAACGCTGCGACCTGCTGATGCGGCGGCTCTG 780  
 Db 720 TCGGCTCAGAGCTCGAGGCTGCGTCAACGCTGCGACCTGCTGATGCGGCGGCTCTG 779  
 QY 781 GTGCGAGGCGCCAAAGCJCACAAATAGTCTGCAATTCATTTGCGCGCATATGAACCA 840  
 Db 780 GTGCGAGGCGCCAAAGCJCACAAATAGTCTGCAATTCATTTGCGCGCATATGAACCA 839  
 QY 841 GGTGCGTACTGTGATATATACCATGACACAGGCGGCTGTTTCGAAGGCTCAGACCG 900  
 Db 840 GGTGCGTACTGTGATATATAGCATGACACAGGCGGCTGTTTCGAAGGCTCAGACCG 899  
 QY 901 ACACGCTACGACGACGCGGAGCTTGGCGTGACAGACGCGTGTCTTACTGCGTGGGAGAC 960  
 Db 900 ACACGCTACGACGACGCGGAGCTTGGCGTGACAGACGCGTGTCTTACTGCGTGGGAGAC 959  
 QY 961 ATGCGCGCTGCGTGGCGAAGACGTCGACCTACGCGCTGACCAACGCGAGATGCGTAT 1020  
 Db 960 ATGCGCGCTGCGTGGCGAAGACGTCGACCTACGCGCTGACCAACGCGAGATGCGTAT 1019  
 QY 1021 GTGCTGAGCTTGGCGACGATAGGCTGGCGGCGGCGGCTGGATCCGCGACTAGCC 1080  
 Db 1020 GTGCTGAGCTTGGCGACGATAGGCTGGCGGCGGCGGCTGGATCCGCGACTAGCC 1079  
 QY 1140 GTGCGCTTACCGACGCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
 Db 1140 GTGCGCTTACCGACGCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1199  
 QY 1201 CACAGCTGGGAGTAAGGAGGAGCATGA 1228  
 Db 1200 CACAGCTGGGAGTAAGGAGGAGCATGA 1227

## RESULT 15

LOCUS AB8750 1229 bp DNA PAT 22-JAN-2000

DEFINITION Sequence 7 from Patent WO9832862.  
 ACCESSION A89750  
 VERSION A89750.1 GI:6738284  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1229)  
 AUTHORS Flohe, I. and Slinn, M.  
 TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
 JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE LEOPOLD (DE); SLINN MAHAVIR (DE)  
 FEATURES  
 source location/Qualifiers  
 1..1229  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 236 a 391 c 382 g 219 t 1 others  
 ORIGIN

Query Match 98.7%; Score 1211.8; DB 5; Length 1229;  
 Best Local Similarity 99.7%; Pred. No. 9.3e-167;  
 Matches 1224; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 ATCTTGAGATTATCGAATCTTCTTATCTAGAGC.ACAGTACGAGAGGGTAATC 60  
 DB 1 ATCTTGAGATTATCGAATCTTCTTATCTAGAGCAGTACGAGAGGGTAATC 60  
 QY 61 ATGGCGCTCGATTATCGACGAGACCAAAACACGATTCGGGTGGCCATCACCCG 120  
 DB 61 ATGGCGCTCGATTATCGACGAGACCAAAACACGATTCGGGTGGCCATCACCCG 120  
 QY 121 GCGGCGCTCGGAGCACTAACCCCTGCTGCGCATGAGTGTCTATCCAGGCAAGTCCGGA 180  
 DB 121 GCGGCGCTCGGAGCACTAACCCCTGCTGCGCATGAGTGTCTATCCAGGCAAGTCCGGA 180  
 QY 121 GCGGCGCTCGGAGCACTAACCCCTGCTGCGCATGAGTGTCTATCCAGGCAAGTCCGGA 180  
 DB 121 GCGGCGCTCGGAGCACTAACCCCTGCTGCGCATGAGTGTCTATCCAGGCAAGTCCGGA 180  
 QY 181 GAGGCGCTCGGAGCACTAACCCCTGCTGCGCATGAGTGTCTATCCAGGCAAGTCCGGA 240  
 DB 181 GAGGCGCTCGGAGCACTAACCCCTGCTGCGCATGAGTGTCTATCCAGGCAAGTCCGGA 240  
 QY 241 GCCGACGAGTGTGGGCGGAGCGCTGATTTATGCTCAAGAGTCAAAAGAACGATAGCGGC 300  
 DB 241 GCCGACGAGTGTGGGCGGAGCGCTGATTTATGCTCAAGAGTCAAAAGAACGATAGCGGC 300  
 QY 301 GAATACGCGCGCTGCGACAGCGGAGATCTGTTCAGTCTTGTGCAATTTGGCGCGCTCA 360  
 DB 301 GAATACGCGCGCTGCGACAGCGGAGATCTGTTCAGTCTTGTGCAATTTGGCGCGCTCA 360  
 QY 301 GAATACGCGCGCTGCGACAGCGGAGATCTGTTCAGTCTTGTGCAATTTGGCGCGCTCA 360  
 DB 301 GAATACGCGCGCTGCGACAGCGGAGATCTGTTCAGTCTTGTGCAATTTGGCGCGCTCA 360  
 QY 361 CGTCTTGACGAGTGTGGTGTGATTCGCGACACGATTCGCTACGAGACCGTC 420  
 DB 361 CGTCTTGACGAGTGTGGTGTGATTCGCGACACGATTCGCTACGAGACCGTC 420  
 QY 421 CAGACCGCGAGCGGAGCACTACCCCTGCTTCCCGGATGAGCGAAGTCCGCGTCACTC 480  
 DB 421 CAGACCGCGAGCGGAGCACTACCCCTGCTTCCCGGATGAGCGAAGTCCGCGTCACTC 480  
 QY 420 CAGACCGCGAGCGGAGCACTACCCCTGCTTCCCGGATGAGCGAAGTCCGCGTCACTC 479  
 DB 420 CAGACCGCGAGCGGAGCACTACCCCTGCTTCCCGGATGAGCGAAGTCCGCGTCACTC 479  
 QY 481 GCGGCGCGAGTGTGGGCGTTCACACTGATGGAACCCAAAGGGGCGCGGCTGCTGATG 540  
 DB 481 GCGGCGCGAGTGTGGGCGTTCACACTGATGGAACCCAAAGGGGCGCGGCTGCTGATG 540  
 QY 480 GCGGCGCGAGTGTGGGCGTTCACACTGATGGAACCCAAAGGGGCGCGGCTGCTGATG 539  
 DB 480 GCGGCGCGAGTGTGGGCGTTCACACTGATGGAACCCAAAGGGGCGCGGCTGCTGATG 539  
 QY 541 GCGGCGCGAGTGTGGGCGTTCACACTGATGGAACCCAAAGGGGCGCGGCTGCTGATG 600  
 DB 541 GCGGCGCGAGTGTGGGCGTTCACACTGATGGAACCCAAAGGGGCGCGGCTGCTGATG 600  
 QY 540 GCGGCGCGAGTGTGGGCGTTCACACTGATGGAACCCAAAGGGGCGCGGCTGCTGATG 599  
 DB 540 GCGGCGCGAGTGTGGGCGTTCACACTGATGGAACCCAAAGGGGCGCGGCTGCTGATG 599  
 QY 601 TACAACGAGCGCGATGCGCAAGCGCATGGCGGAGCGGATGAGTCTAGACATCAAC 660  
 DB 601 TACAACGAGCGCGATGCGCAAGCGCATGGCGGAGCGGATGAGTCTAGACATCAAC 660  
 QY 600 TACAACGAGCGCGATGCGCAAGCGCATGGCGGAGCGGATGAGTCTAGACATCAAC 659  
 DB 600 TACAACGAGCGCGATGCGCAAGCGCATGGCGGAGCGGATGAGTCTAGACATCAAC 659  
 QY 661 ATGCAACAATCTCGCAACTCGAGCCGAGTCTGCGGCGCGATCACAACCTCGCTACTCA 720  
 DB 661 ATGCAACAATCTCGCAACTCGAGCCGAGTCTGCGGCGCGATCACAACCTCGCTACTCA 720  
 QY 660 ATGCAACAATCTCGCAACTCGAGCCGAGTCTGCGG CGGATCACAACCTCGCTACTCA 719  
 DB 660 ATGCAACAATCTCGCAACTCGAGCCGAGTCTGCGG CGGATCACAACCTCGCTACTCA 719  
 QY 721 TCGGCTACGAGCTCGAGGCTGCGCTCAAAAGTCCGAGCTGCTGATTTGGGCGCTCTG 780  
 DB 721 TCGGCTACGAGCTCGAGGCTGCGCTCAAAAGTCCGAGCTGCTGATTTGGGCGCTCTG 780

DB 720 TCGGCTACGAGCTCGAGGCTGCGCTCAAAAGTCCGAGCTGCTGATTTGGGCGCTCTG 779  
 QY 781 GTCCAGAGCGCGCAAGGCGCAACCAATTAATCTGCAATTAATCTGCGCATATGAACCA 840  
 DB 780 GTCCAGAGCGCGCAAGGCGCAACCAATTAATCTGCAATTAATCTGCGCATATGAACCA 839  
 QY 841 GGTGCGTACTGTTGATATAGCATGACAGGCGGCTGTTTGAAGGCTACAGACCG 900  
 DB 840 GGTGCGTACTGTTGATATAGCATGACAGGCGGCTGTTTGAAGGCTACAGACCG 899  
 QY 901 ACCACCTACGACACCGCAAGCTTCCCGCTGACAGCAACGCTGTTTACTGCTGCGGAC 960  
 DB 900 ACCACCTACGACACCGCAAGCTTCCCGCTGACAGCAACGCTGTTTACTGCTGCGGAC 959  
 QY 961 ATGCCGCGCTGCTGCGCAAGAGCTGACCTAGCGGCTGACCAACGCGAGATGCGGTAT 1020  
 DB 960 ATGCCGCGCTGCTGCGCAAGAGCTGACCTAGCGGCTGACCAACGCGAGATGCGGTAT 1019  
 QY 1021 GTGCTGAGCTTCCGAGCATGCTGCGGCGGCGCTGCGGCTGCAATCCGCACTAGCC 1080  
 DB 1020 GTGCTGAGCTTCCGAGCATGCTGCGGCGGCGCTGCGGCTGCAATCCGCACTAGCC 1079  
 QY 1081 AAAGGCTCTTGCAGCGCAGAAAGGCGGCTTACTGTCCGAAGGAGTGGCCAGCAGCTGGGG 1140  
 DB 1080 AAAGGCTCTTGCAGCGCAGAAAGGCGGCTTACTGTCCGAAGGAGTGGCCAGCAGCTGGGG 1139  
 QY 1141 GTGCGCTTACCGAGCCCGCCAGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
 DB 1140 GTGCGCTTACCGAGCCCGCCAGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1199  
 QY 1201 CACAGCTGGGAGTAAGGAGGAGATGA 1228  
 DB 1200 CACAGCTGGGAGTAAGGAGGAGATGA 1227

Search completed: June 22, 2000, 14:53:39  
 Job time: 17634 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:07:26 ; Search time 458.59 Seconds  
(without alignments)  
669,958 Million cell updates/sec

Title: US-09-362-485-5

Perfect score: 1228  
Sequence: 1 ATCTTCAGATTATATCGAAC.....GGGAGTACGAGGAGCATCA 1228

Scoring table: IDENTITY-MUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 23170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance - you have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1210.4	98.6	1245	1 V49626	Mycobacterium tube
2	1210.4	98.6	1260	1 V49510	Mycobacterium sp.
3	1210.4	98.6	1260	1 V49625	Mycobacterium tube
4	677.6	38.1	682	1 V49511	Mycobacterium mari
5	277.2	22.6	1125	1 N91423	Sequence of heat-r
6	212.4	17.3	28171	1 V52155	Streptococcus pneu
7	178.4	14.5	9280	1 V74442	Streptococcus aur
8	99.2	5.6	544	1 T67971	H. pylori membrane
9	69.2	4.7	31096	1 V43039	Streptococcus pneu
10	57.6	4.1	1074	1 V74370	Streptococcus aur
11	50.8	4.1	390	1 Q21833	Randomising oligon
12	50.8	4.1	390	1 Q36859	PCR primer for 5'
13	50.4	4.1	1413	1 X34205	Mycobacterium spec
14	50.4	4.1	1722	1 X34206	Mycobacterium spec
15	50.4	4.1	15872	1 T68715	Streptomyces venez
16	49	4.0	1269	1 Q03065	Sequence encoding
17	47.6	3.9	3966	1 T93610	Mycobacterium tube
18	47.6	3.9	28598	1 T06769	Sorangium cellulos
19	47.6	3.9	28958	1 T89956	Sorangium cellulos
20	47.6	3.9	49377	1 V05287	The soraphen biosy
21	46.4	3.8	985	1 V44439	Mycobacterium tube
22	46.4	3.8	985	1 V64548	M. tuberculosis im
23	46	3.7	1620	1 Q22482	groEL-1 gene codin
24	46	3.7	1582	1 T73117	Actinoplanes sp. a
25	46	3.7	2582	1 Q22485	groEL-1 gene. Reco
26	45	3.7	17955	1 V56642	Actinoplanes sp. a
27	45.4	3.7	833	1 Q64201	snab gene encoding
28	45.4	3.7	5392	1 Q64201	Sequence compri
29	42.4	3.5	1833	1 V84066	snab gene encoding
30	42.4	3.5	2185	1 V84066	Clone p5-4 encodin
31	42.4	3.5	2185	1 Q99365	S. lividans protea
32	42.4	3.5	24379	1 T93095	Streptomyces fireo
33	42.4	3.5	24379	1 V25925	Streptomyces roseo
34	42.2	3.4	329	1 V44425	Mycobacterium tube

35	42.2	3.4	329	1 V64534	M. tuberculosis im
36	42	3.4	3946	1 T93610	Mycobacterium tube
37	42	3.4	4018	1 Q63879	Polyhydroxyalkanoa
38	41.8	3.4	882	1 V44403	Mycobacterium tube
39	41.8	3.4	882	1 V64512	M. tuberculosis im
40	41.6	3.4	12036	1 Q04668	FHA structural gen
41	41.2	3.4	2414	1 Q05926	Sequence encoding
42	41	3.3	29879	1 Q46806	eryA region of S.
43	40.8	3.3	2151	1 Q76252	HSV-2 protease, IC
44	40.8	3.3	2151	1 Q76261	HSV-2 protease/ICP
45	40.8	3.3	2472	1 Q84671	HSV-2 UL26 gene. N

## ALIGNMENTS

RESULT 1	
V49626	1
ID	V49626 standard; DNA; 1245 BP.
AC	V49626:
DT	20-NOV-1998 (first entry)
DE	Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW	ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS	Mycobacterium tuberculosis.
PN	W09836089-A2.
PD	20-AUG-1998.
PF	29-JAN-1998: E00483.
PR	29-JAN-1997: EP-101338.
PA	(FLOH/) FLOHE L.
PI	Flohe L, Butter B, Kolk A, Singh M;
DR	WPI; 98-457123/39.
PT	Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity.
PT	- useful for, e.g. for diagnosis, differentiation of strains,
PT	monitoring vaccination and identification of mycobacterial
PT	inhibitors
PS	Disclosure; Fig 3.19; 55pp; German.
CC	The Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC	the production of kits for diagnosing tuberculosis (TB) and other
CC	mycobacterial infections in humans or animals. Kits are used for direct
CC	diagnosis of TB on clinical samples (e.g. body fluids) and can
CC	differentiate between pathogenic and non-virulent strains, e.g. for
CC	identifying pathogens (of the M. tuberculosis (M. t.) complex. The kit may
CC	also be used to identify substances that inhibit mycobacteria, for
CC	combating epidemics and for vaccination follow-up. Oligonucleotides
CC	derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC	also for culture confirmation of isolated strains and for chromosome
CC	fingerprinting to detect/differentiate between mycobacteria, and for
CC	L-alanine-specific biotransformation reactions. AlaDH is an early
CC	antigen, present extracellularly after only a few days of growth, making
CC	it an ideal drug target.
SQ	Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 98.6%; Score 1210.4; DB 1; Length 1245;  
Best Local Similarity 99.4%; Pred. No. 8e+244;  
Matches 1227; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY	1	ATCTTCAGATTATATCGAACCTTCTTACTAGACGTACAGTATCGAGGGGTATC 60
DB	1	ATCTTCAGATTATATCGAACCTTCTTACTAGACGTACAGTATCGAGGGGTATC 60
QY	61	ATGCGGTGCGATTTCGACCGACGACCAAAACACG-----AATCCGGGTGGCCATC 114
DB	61	ATGCGGTGCGATTTCGACCGACGACCAAAACACG-----AATCCGGGTGGCCATC 114
QY	121	ATGCGGTGCGATTTCGACCGACGACCAAAACACGAAATTCATTCGGGTGGCCATC 120
DB	121	ATGCGGTGCGATTTCGACCGACGACCAAAACACGAAATTCATTCGGGTGGCCATC 120
QY	115	ACCCGCGCGGCGTGGCGACTACCCGTCGTGGCGATGAGGTCTCATCCAGGACGT 174
DB	115	ACCCGCGCGGCGTGGCGACTACCCGTCGTGGCGATGAGGTCTCATCCAGGACGT 174
QY	175	GCGGAGAGGCGTGGCGATTACCGACGCGGATTTCAAGCGCGGCGGCAATGGTC 234
DB	175	GCGGAGAGGCGTGGCGATTACCGACGCGGATTTCAAGCGCGGCGGCAATGGTC 234
QY	181	GCGGAGAGGCGTGGCGATTACCGACGCGGATTTCAAGCGCGGCGGCAATGGTC 240
DB	181	GCGGAGAGGCGTGGCGATTACCGACGCGGATTTCAAGCGCGGCGGCAATGGTC 240

```

QY 235 GGCACGCGCCGACAGTGTGGCCGACGCTGATTATTGCTCAAGGTCAAAAGACGATA 294
   |||||||
DB 241 GGCACCGCCGACAGTGTGGCCGACGCTGATTATTGCTCAAGGTCAAAAGACGATA 300
   |||||||
QY 295 GCGGCGGAATACGGCCGCTGCGACACGGGCGAGATCTTGTTCAGTCTTGTGATTTGGCC 354
   |||||||
DB 301 GCGGCGGAATACGGCCGCTGCGACACGGGCGAGATCTTGTTCAGTCTTGTGATTTGGCC 360
   |||||||
QY 355 GCGTACGCTGTGACGCGATGCTGTTGGATTCGCGACACGATTCATTCCTTACGAG 414
   |||||||
DB 361 GCGTACGCTGTGACGCGATGCTGTTGGATTCGCGACACGATTCATTCCTTACGAG 420
   |||||||
QY 415 ACCGTCCAGACCGCCGACGCGACCTACCCCTGCTTGGCCCATGACGCAAGTCCCGGT 474
   |||||||
DB 421 ACCGTCCAGACCGCCGACGCGACCTACCCCTGCTTGGCCCATGACGCAAGTCCCGGT 480
   |||||||
QY 475 CGACTCGCCGCGCGCTTACGACCTGATCGCAACCAAGGGGGCGCGGTG 534
   |||||||
DB 481 CGACTCGCCGCGCGCTTACGACCTGATCGCAACCAAGGGGGCGCGGTG 540
   |||||||
QY 535 CTGATGGCGGGGTGCGCGCGCTGCAACCGCGCGACGCTGTGTATCGCGCCGCGAC 594
   |||||||
DB 541 CTGATGGCGGGGTGCGCGCGCTGCAACCGCGCGACGCTGTGTATCGCGCCGCGAC 600
   |||||||
QY 595 GCGGCTTAACGCAACCGCGCATCGCAACGGCGATGGCGCAACCGTTACGGTTCTAGAC 654
   |||||||
DB 601 GCGGCTTAACGCAACCGCGCATCGCAACGGCGATGGCGCAACCGTTACGGTTCTAGAC 660
   |||||||
QY 655 ATCAACATGACAACTTGGGCACTCGACGCGGAGTTCTGCGCGGATTCACACTCGC 714
   |||||||
DB 661 ATCAACATGACAACTTGGGCACTCGACGCGGAGTTCTGCGCGGATTCACACTCGC 720
   |||||||
QY 715 TACTCATCGCCTACGACCTCGAGGCTGCGCTCAAAAGTCCGACCTGTGATGGGCGC 774
   |||||||
DB 721 TACTCATCGCCTACGACCTCGAGGCTGCGCTCAAAAGTCCGACCTGTGATGGGCGC 780
   |||||||
QY 775 GTCCTGCTGCGAGGCGCGCAAGGCAATTAATGCTCAATTAATTCCTTCCGCGCATG 834
   |||||||
DB 781 GTCCTGCTGCGAGGCGCGCAAGGCAATTAATGCTCAATTAATTCCTTCCGCGCATG 840
   |||||||
QY 835 AAACAGGTGCGGTGCTGATATAGCCATGACGACGAGGCGGCTGTTTCGAAGCTCA 894
   |||||||
DB 841 AAACAGGTGCGGTGCTGATATAGCCATGACGACGAGGCGGCTGTTTCGAAGCTCA 900
   |||||||
QY 895 CGACCGACACCTAGACACCAACCGGCTGCGCGTGCAGACAGACGCTGTTTACTGCGTG 954
   |||||||
DB 901 CGACCGACACCTAGACACCAACCGGCTGCGCGTGCAGACACGCTGTTTACTGCGTG 960
   |||||||
QY 955 GCGAACATGCGCGCTGCGGTGCGGAAGAGCTGACCTAGCGCGTGCACCAACGCGAGATG 1014
   |||||||
DB 961 GCGAACATGCGCGCTGCGGTGCGGAAGAGCTGACCTAGCGCGTGCACCAACGCGAGATG 1020
   |||||||
QY 1015 CCGTATGTGCTGACGCTTGGCCGACATGCTGCGGGGGCGGCTGCGGATTCGCGCA 1074
   |||||||
DB 1021 CCGTATGTGCTGACGCTTGGCCGACATGCTGCGGGGGCGGCTGCGGATTCGCGCA 1080
   |||||||
QY 1075 CTAGCAAAAGGCTTTCGACGACGGAAGGGGCTTACTGTCGAAGGGGTCGACGAC 1134
   |||||||
DB 1081 CTAGCAAAAGGCTTTCGACGACGGAAGGGGCTTACTGTCGAAGGGGTCGACGAC 1140
   |||||||
QY 1135 CTGCGGGGTGCGGTTCACGACGAGCGCGCAGCGTCTGCGCTGACCTGCGCGCTGTAC 1194
   |||||||
DB 1141 CTGCGGGGTGCGGTTCACGACGAGCGCGCAGCGTCTGCGCTGACCTGCGCGCTGTAC 1200
   |||||||
QY 1195 GCGGACGACAGTCTGCGAGTAAGGAAGGATGA 1228
   |||||||
DB 1201 GCGGACGACAGTCTGCGAGTAAGGAAGGATGA 1234
   |||||||

```

RESULT 2  
 V49510  
 ID V49510 standard; DNA; 1260 BP.  
 AC V49510;

```

DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AladH DNA.
KW Alanine dehydrogenase; AladH; ADH; diagnosis: tuberculosis; pathogen;
   swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN M09832862-A2.
PD 30-JUL-1998.
PE 29-JAN-1998; E00484.
PF 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L. Kolk A. Singh M.
PI Fione L, Butler B, Kolk A. Singh M.
DR WPI: 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
   - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
   for bio-transformation
PS Disclosure, Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
   from a Mycobacterium sp. This protein is used to diagnose tuberculosis
   and other mycobacterial infections (including 'swimmers' disease', caused
   by M. marinum, a fish pathogen) in humans or animals. The protein can
   also be used for control of epidemics and for vaccination, to screen for
   agents with anti-mycobacterial activity, and in bio-transformations that
   are specific for L-alanine. Also mycobacteria can be identified by
   CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
   CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 98.6%; Score 1210.4; DB 1; Length 1260;
Best Local Similarity 99.4%; Pred. No. 8e-244;
Matches 1227; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

QY 1 ATCTTGCAATTAATGCACTTCTTATGATGAGCGTACAGTACGAGGAGGTATC 60
   |||||||
DB 16 ATCTTGCAATTAATGCACTTCTTATGATGAGCGTACAGTACGAGGAGGTATC 75
   |||||||
QY 61 ATGGCGGTGCGTATTCGACCGGACCAAAACAGC-----AATTCGGGTGGCCATC 114
   |||||||
DB 76 ATGGCGGTGCGTATTCGACCGGACCAAAACAGTATTCGAATTCGGGTGGCCATC 135
   |||||||
QY 115 ACCCGCGCGCGCGTGGCGAATACCGCGTGTGGCGCATGAGTGTCTATCAGGACGAT 174
   |||||||
DB 136 ACCCGCGCGCGCGTGGCGAATACCGCGTGTGGCGCATGAGTGTCTATCAGGACGAT 195
   |||||||
QY 175 GCGGAGAGGCGTGGCTATACCGACGCGGATTTCAAGCGCGGACGCGCAACTGTC 234
   |||||||
DB 196 GCGGAGAGGCGTGGCTATACCGACGCGGATTTCAAGCGCGGACGCGCAACTGTC 255
   |||||||
QY 235 GGCACCGCGCGACAGTGTGGCGGACGCTGATTTATGCTCAAGGTCAAAGAACGATA 294
   |||||||
DB 256 GGCACCGCGCGACAGTGTGGCGGACGCTGATTTATGCTCAAGGTCAAAGAACGATA 315
   |||||||
QY 295 GCGGCGGAATACGGCCGCTGCGACGCGGAGATCTTGTTCAGCTTCTTGCATTTGGCC 354
   |||||||
DB 316 GCGGCGGAATACGGCCGCTGCGACGCGGAGATCTTGTTCAGCTTCTTGCATTTGGCC 375
   |||||||
QY 355 GCGTACGCTGTGACGCGATGCTGTTGGATTCGCGACACGCTCAATTTGCTTACGAG 414
   |||||||
DB 376 GCGTACGCTGTGACGCGATGCTGTTGGATTCGCGACACGCTCAATTTGCTTACGAG 435
   |||||||
QY 415 ACCGTCCAGACCGCCGACGCGGACCTACCCCTGCTTGGCCCATGAGGAGGAGTGGCCGT 474
   |||||||
DB 436 ACCGTCCAGACCGCCGACGCGGACCTACCCCTGCTTGGCCCATGAGGAGGAGTGGCCGT 495
   |||||||
QY 475 CGACTCGCCGCGCGCTTACGACCTGATCGCAACCAAGGGGGCGCGGTG 534
   |||||||
DB 496 CGACTCGCCGCGCGCTTACGACCTGATCGCAACCAAGGGGGCGCGGTG 555
   |||||||
QY 535 CTGATGGCGGGGTGCGCGCGCTGCAACGGCGCGAGTGTGGATCGGCGCGGAC 594
   |||||||
DB 556 CTGATGGCGGGGTGCGCGCGCTGCAACGGCGCGAGTGTGGATCGGCGCGGAC 615
   |||||||

```



QY 595 GCCGGCTACACGACGACCCGATCGCAACGCGATGGCGCGACCGTTACGGTTCTAGAC 654  
 DB 616 GCCGGCTACACGACGACCCGATCGCAACGCGATGGCGCGACCGTTACGGTTCTAGAC 675  
 QY 655 ATCAACATGACAACTTGGCACTGACGCGGATCTGGGGCGGATCCACACTGCG 714  
 DB 676 ATCAACATGACAACTTGGCACTGACGCGGATCTGGGGCGGATCCACACTGCG 735  
 QY 715 TACTCATCGGCTACGAGCTCGAGGGTGGCGCAACGTCGCGACCTGGTATGGGGCC 774  
 DB 736 TACTCATCGGCTACGAGCTCGAGGGTGGCGCAACGTCGCGACCTGGTATGGGGCC 795  
 QY 775 GTCTGTGTCGACGCGCGCAAGGACCCAAATTAGTCTGCAATTCTGCGCGCATATG 834  
 DB 796 GTCTGTGTCGACGCGCGCAAGGACCCAAATTAGTCTGCAATTCTGCGCGCATATG 855  
 QY 835 AAACGAGTGGCTACTGGTATATAGCATCGACGAGGGGCTGTTCCAAAGGCTCA 894  
 DB 856 AAACGAGTGGCTACTGGTATATAGCATCGACGAGGGGCTGTTCCAAAGGCTCA 915  
 QY 895 CGACGACGACCTACGACGACCCGATCGCGCTGCAACGACGCTGTTTACTGCGTG 954  
 DB 916 CGACGACGACCTACGACGACCCGATCGCGCTGCAACGACGCTGTTTACTGCGTG 975  
 QY 955 GCGACATGCGCGCTGCGTGGTGGCAAGAGCTGACCTACGCGCTGACCAACGCGCATG 1014  
 DB 976 GCGACATGCGCGCTGCGTGGTGGCAAGAGCTGACCTACGCGCTGACCAACGCGCATG 1035  
 QY 1015 CCTATGCTGCTGAGCTGGCGCAACATGGCTGGCGGGGCGTGGCGGATCCGGCA 1074  
 DB 1036 CCTATGCTGCTGAGCTGGCGCAACATGGCTGGCGGGGCGTGGCGGATCCGGCA 1095  
 QY 1075 CTAGCCAAAGGCTTTCGACGACGAGAGGGGCTTACTGTCCGACGGGTGGCGACGAC 1134  
 DB 1096 CTAGCCAAAGGCTTTCGACGACGAGAGGGGCTTACTGTCCGACGGGTGGCGACGAC 1155  
 QY 1135 CTGGGGGCTGCTTACCGACGACCGCGCGAGCGTGTGCTGCTGCTGCTGCTGCTTAC 1194  
 DB 1156 CTGGGGGCTGCTTACCGACGACCGCGCGAGCGTGTGCTGCTGCTGCTGCTTAC 1215  
 QY 1195 GCCGAGACGCTCGGAGTAAGGAAGGATGA 1228  
 DB 1216 GCCGAGACGCTCGGAGTAAGGAAGGATGA 1249

RESULT 3  
 ID V49625  
 AC V49625:  
 DT 20-NOV-1998 (first entry)  
 DE Mycobacterium tuberculosis Alamine dehydrogenase.  
 KW ss: Alamine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.  
 OS Mycobacterium tuberculosis.  
 PN M09836089-A2.  
 PD 20-JAN-1998.  
 PR 28-JAN-1998; E00483.  
 PR 29-JAN-1997; EP-101338.  
 PA (FLOH/) FLOHE L.  
 PI Flohe L, Hutter B, Kolk A, Singh M;  
 DR WPI: 98-457123/39.  
 PT Diagnosis of, e.g. tuberculosis from alamine dehydrogenase activity  
 PT - useful for, e.g. for diagnosis, differentiation of strains,  
 PT monitoring vaccination and identification of mycobacterial  
 PT inhibitors  
 PS Claim 13; Fig 2.3; 55pp; German.  
 CC The Mycobacterium tuberculosis alamine dehydrogenase (AlaDH) is in  
 CC the production of kits for diagnosing tuberculosis (TB) and other  
 CC mycobacterial infections in humans or animals. Kits are used for direct  
 CC diagnosis of TB on clinical samples (e.g. body fluids) and can  
 CC differentiate between pathogenic and non-virulent strains, e.g. for  
 CC identifying pathogens of the M. tuberculosis (M. t.) complex. The kit may  
 CC also be used to identify substances that inhibit mycobacteria, for  
 CC combatting epidemics and for vaccination follow-up. Oligonucleotides

CC derived from AlaDH are used similarly in diagnostic hybridisation tests,  
 CC also for culture confirmation of isolated strains and for chromosome  
 CC fingerprinting to detect/differentiate between mycobacteria, and for  
 CC L-alanine-specific biotransformation reactions. Mycobacteria, and for  
 CC antigen, present extracellularly after only a few days of growth, making  
 CC it an ideal drug target.  
 SO Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 98.6%; Score 1210.4; DB 1; Length 1260;  
 Best Local Similarity 99.4%; Pred. No. 8e-244;  
 Matches 1227; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

QY 1 ATCTTGACGATTATGACACTTCTTCTACGAGCTACGATGACGAGGGGTATTC 60  
 DB 16 ATCTTGACGATTATGACACTTCTTCTACGAGCTACGATGACGAGGGGTATTC 75  
 QY 61 ATGCGCTGCGTATTCGACGAGACCAAAACAAG-----AATTGCGGGTGGCATC 114  
 DB 76 ATGCGCTGCGTATTCGACGAGACCAAAACAAGATTTCGATTCGCGGTGGCATC 135  
 QY 115 ACCCGCGCGCGCTGCGGACGACCTACCGCTGCGGACGATGAGGTGCTATCGAGCAGT 174  
 DB 136 ACCCGCGCGCGCTGCGGACGACCTACCGCTGCGGACGATGAGGTGCTATCGAGCAGT 195  
 QY 175 GCGGAGAGAGGCTCGGCTATACCGACGCGGATTTCAAGGGCGGCGCCACCTGGTC 234  
 DB 196 GCGGAGAGAGGCTCGGCTATACCGACGCGGATTTCAAGGGCGGCGCCACCTGGTC 255  
 QY 235 GCGGAGAGAGGCTCGGCTATACCGACGCGGATTTCAAGGGCGGCGCCACCTGGTC 294  
 DB 256 GCGGAGAGAGGCTCGGCTATACCGACGCGGATTTCAAGGGCGGCGCCACCTGGTC 315  
 QY 295 GCGGAGAGAGGCTCGGCTATACCGACGCGGATTTCAAGGGCGGCGCCACCTGGTC 354  
 DB 316 GCGGAGAGAGGCTCGGCTATACCGACGCGGATTTCAAGGGCGGCGCCACCTGGTC 375  
 QY 355 GCGTCACTGCTTTCGACGAGTCTGTTGATTCGCGGACGACGATCTGCTGATTCAGAG 414  
 DB 376 GCGTCACTGCTTTCGACGAGTCTGTTGATTCGCGGACGACGATCTGCTGATTCAGAG 435  
 QY 415 ACCGTGACAGCGCGGACGAGGCGGACGATCTGTTGACGTTCTGATTCAGAGTTCAGAG 474  
 DB 436 ACCGTGACAGCGCGGACGAGGCGGACGATCTGTTGACGTTCTGATTCAGAGTTCAGAG 495  
 QY 475 CGACTGCGCGCGGAGTGGCTTACCACTGATTCGAGCAACCAAGGGGCGCGGTGTG 534  
 DB 496 CGACTGCGCGCGGAGTGGCTTACCACTGATTCGAGCAACCAAGGGGCGCGGTGTG 555  
 QY 535 CTGATGGGCGGGGTGCGCGCTGCAACGCGGACGATGATGGCGCGCGGACG 594  
 DB 556 CTGATGGGCGGGGTGCGCGCTGCAACGCGGACGATGATGGCGCGCGGACG 615  
 QY 595 GCCGGCTACACGACGACCCGATCGCAACGCGATGGCGCGACCGTTACGGTTCTAGAC 654  
 DB 616 GCCGGCTACACGACGACCCGATCGCAACGCGATGGCGCGACCGTTACGGTTCTAGAC 675  
 QY 655 ATCAACATGACAACTTGGCACTGACGCGGATCTGGGGCGGATCCACACTGCG 714  
 DB 676 ATCAACATGACAACTTGGCACTGACGCGGATCTGGGGCGGATCCACACTGCG 735  
 QY 715 TACTCATCGGCTACGAGCTCGAGGGTGGCGCAACGTCGCGACCTGGTATGGGGCC 774  
 DB 736 TACTCATCGGCTACGAGCTCGAGGGTGGCGCAACGTCGCGACCTGGTATGGGGCC 795  
 QY 775 GTCTGTGTCGACGCGCGCAAGGACCCAAATTAGTCTGCAATTCTGCGCGCATATG 834  
 DB 796 GTCTGTGTCGACGCGCGCAAGGACCCAAATTAGTCTGCAATTCTGCGCGCATATG 855  
 QY 835 AAACGAGTGGCTACTGGTATATAGCATCGACGAGGGGCTGTTCCAAAGGCTCA 894  
 DB 856 AAACGAGTGGCTACTGGTATATAGCATCGACGAGGGGCTGTTCCAAAGGCTCA 915

Qy	895	CGCCGACACACCTACGACACACCCGAGCTTGCGCGTACGACGACCTGTTTACGCGTG	954
Db	916	CGACCGACACCTACGACACACCCGACGTTGCGCGTACGACGACCGCTTTTACGCGTG	975
Qy	955	GCGAACAATGCCCGCGCTGCGTGCCGAGAGAGCTGCACCTACGCGCTGACCAAGCCGACATG	1014
Db	976	GCGAACAATGCCCGCGCTTGCGTGCCGAGAGAGCTGCACCTACGCGCTACCAAGCCGACGATG	1035
Qy	1015	CCGATATGCGTGCAGCTTGTCCGACCAATGGCTGGCGCGCGCGCTGCCTGCGTGCATCCGACA	1074
Db	1036	CCGATATGCGTGCAGAGCTTGGCGACCAATGGCTGGCGCGCGCGCTGCATCCGACA	1095
Qy	1075	CTAACCACAAAGTCTTTTGCAGACGACAGAAAGGGCGTTACTTCGAAAGGGTGGCCACCGAC	1134
Db	1096	CTAACCACAAAGTCTTTTGCAGACGACAGAAAGGGCGTTACTGTCGAAAGGGTGGCCACCGAC	1155
Qy	1135	CTGGGGGTCGCCGTTACACCGAGCCCGCAGCGTGTGAGCTCTGGCGCGCTGTTAC	1194
Db	1156	CTGGGGGTCGCCGTTACACCGAGCCCGCAGCGTGTGAGCTCTGGCGCGCTGTTAC	1215
Qy	1195	GCGACGACACGTCGCGGAGTAGTAAGGAAGCATGA	1228
Db	1216	GCGACGACACGTCGCGGAGTAGTAAGGAAGCATGA	1249

RESULT	4
ID	V49511
AC	V49511; standard; DNA; 682 BP.
DT	20-OCT-1998 (first entry)
DE	Mycobacterium marinum Mar3 DNA.
KM	Alanine dehydrogenase; Aladh; ADH; diagnosis; tuberculosis; pathogen;
OS	swimmers disease; vaccine; epidemic; infection; identification; ss.
NC	Mycobacterium marmum.
PR	Key
FT	CDS
FT	Location/Qualifiers
FT	1..682
FT	/tag_a
FT	/codon_start=3
FT	/product="Mar3"
FT	/note="Alanine dehydrogenase"
PT	MO9832862-A2.
PD	30-JUL-1998
PF	29-JAN-1998; E00484.
PE	29-JAN-1997; EP-101339.
PI	(FLOH/) FLOHE L.
PA	Flohe L., Rutter B., Kolk A., Singh M:
PB	WPI: 98-427958/36.
DR	P-P5PB: W64481.
PT	Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT	- used for diagnosis of tuberculosis and other mycobacterial
PT	diseases, also for treatment and prevention, for drug screening and
PT	for bio-transformation
PS	Claim 1: Page 34-35; 57pp; German.
CC	This sequence encodes an alanine dehydrogenase (ADH) protein, Mar3
CC	isolated from Mycobacterium marmum. This protein is used to diagnose
CC	tuberculosis and other mycobacterial infections (including 'swimmers'
CC	disease, caused by M. marinum, a fish pathogen) in humans or animals.
CC	The protein can also be used for control of epidemics and for
CC	vaccination, to screen for agents with anti-mycobacterial activity, and
CC	in bio-transformations that are specific for L-alanine. Also mycobacte-
CC	ria can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC	that is secreted early during infection.
SQ	Sequence 682 BP: 105 A; 254 C; 225 G; 98 T;

Query Match	38.1%	Score 467.6;	DB 1;	Length 682;
Best Local Similarity	80.4%	Pred. No. 2.9e-89;		
Matches 548; Conservative	0;	Mismatches 134;	Indels 0;	Gaps
Qy	98	AATTCGGGTGCATCA	●	CCCCGGCGGGCGGGAGAACTAACCGTCGTCGCATAGG 157
Db	1	AATTCGGGTAGCGATCCACCGCGCGGGGTGCGCCCTTGACCAAGCGCGGCACGAG 60		

QY	158	TGCTCATCAGGCGAGGTGGCGAGAGGGGCTTGGCTTTCACCGAGGGGATTCGTAAGGCG	21
Db	61	TGCTGATCCAGGCGCGGTGGCGGAAGAGCTCCGCCATCTCCGAGCGCGAGTCGTAAGCGCG	120
QY	218	CAGGCGCGCACTGTTGGTGGGCAACCGCCGACAGGTGGTGGGCGCGAGTATTGTTCTCA	277
Db	121	CCGGTGGCCCACTGATATAGCACCGCGACCAAGGAGTGGGGGATGGGAGACTCTCTCTCA	180
QY	278	AGGTCAAGAAGCCGATAGCGGCGGGAATACGGCGCGCTCGCAGCAACGGGAGAGTCTTTCA	337
Db	181	AGGTCAAGAAGCCATCGAGTCGAGTCGAGTACGGCGGCGTGGCCCGGGGCGCAACCTCTTTCA	240
QY	338	CGTCTTGCATTTGGCGCGCGTCAAGTGTCTTGCACCGATGGTGGTGTGGATTCCGGCAGCA	397
Db	241	CTTACTGTCACTTGGCGCGCTCTCGCGCCCTCTGCACCGATGCCCTGTGTAAATCCGGGACCA	300
QY	398	CGTCAATTGCTCTACGAGTCGTCGAGACGCGCGAGCGGCACTTACCTCTGTGGCCCGA	457
Db	301	CGTCAATGCTCTACGAGCGGTGCGAGCCCGCGAGCGCGGATTCGCCCTCTGGGCCCA	360
QY	458	TGAGGGAATGGCGCGGTGCACTCCCGCCCGAGGTGGCGCTTACCACTGATCGAACC	517
Db	361	TGAGGAGAGTGGCGCGGCGGCGCTGTCCGCCCAAGTGGGGCTTACCACTATGGGACCC	420
QY	518	AAGGGGCGCGGTGTGCTGATGGGGCGGGTGCCTCGCGCTGGAACCGCGCGAGTGTGG	577
Db	421	ACGGGCGGTGGGGGTGTGATGGGCGCGGTGCCCGCGCTAACCCGTGGCGAGTGTGG	480
QY	578	TGATCGGCGCGCGGCAACCGCGGCTACAAAGCGAGCGCCCATCGCCACCAAGGATGGGCGGA	637
Db	481	TGATGGGCGCGGCGGCAACCGCGGCGGATACAAAGCGCGCGCGGTGGCCCAAGCGATGGGCGGA	540
QY	638	CCGTACGGTCTAGACATCAACATCGACAACCTTGGGCACTGGAGCGCGAGTCTGGG	697
Db	541	TGGTATCCGCTGTGATGTCAACATCAACAAGCTCCGCAATGCAAGCGGAGTGTGGCG	600
QY	698	GCGGATCCACACTGCTACTCATCATGAGCTTACGAGCTTGAAGGTTGCCGTCAAAAGTGGC	757
Db	601	GTCGGGTCCGAGCCCGCTACTGTGTGACCTTCGAACTTGAGAGATGGCGAGATCCACGCGG	660
QY	758	ACCTGTGATTGGGCGCTCCT	779
Db	661	ACATGTATCGGGCGCTCCT	682

RESULT 5  
ID N91423  
AC N91423 standard; DNA; 1125 BP.  
DT 01-FEB-1991 (first entry)  
DE Sequence of heat-resistant alanine dehydrogenase (AH) gene with mol. wt  
KM 2 MD or less  
KW Enzyme; ds.  
OS *Bacillus stearothermophilus* IFO 12550.  
PN J01043194-R.  
PD 15-FEB-1989.  
PF 10-AUG-1987; 200524.  
PR 10-AUG-1987; JP-200524.  
PA (NIRA) Unilika KK.  
DR MPI: 89-096096/13.  
PT Recombinant plasmid for transforming *Escherichia coli* -  
PT obtd. by connecting heat resistant alanine dehydrogenase gene to  
PT vector plasmid  
PS Disclosure; Fig 2, p693; 9pp; Japanese.  
CC A recombinant plasmid conty. heat-resistant alanine dehydrogenase (AH)  
CC gene with mol. wt. 2 or less MD is claimed. Cells transformed with the  
CC vector produce high levels of heat-resistant AH. Transformed *E. coli*  
CC is useful as a clinical inspection reagent.  
SQ Sequence 1125 BP; 246 A; 293 C; 367 G; 217 T;

Query Match	22.68;	Score 277.2;	DB 1;	Length 1125;
Best Local Similarity	56.28;	Pred. No. 1.3e-49;		



```

OY 361 CGTGGTGGACCGATGCGTGTGGATTCCGGACACGTAATTGCTTACGAGACGCTC 420
DB 21414 CAGAGATTAGCAGAGATGCTATGTT-----AACGCAAAAAACAAGTAACCTGTT 21367
OY 421 CAGACCGCGGACGAGGACATACCCCTGCTGGCCCGATGAGGAGGAGTGGCGGTGACTC 480
DB 21366 CGTGACATTCAGAGACAACTACCGGCTCGTCTCTATGAGAGTGGAGTGGATGCTGATG 21307
OY 481 GCCGCCGAGGTGGCGCTTACCACTGATGACCAACCAAGGGGCGCGGTGCTGATG 540
DB 21306 GGTGTTCAATGAGAGCTACTCTCTACTAAGCAAGTGTGGCTGTGTTTACTT 21247
OY 541 GCGGGGGTGGCGGCTGCAACGCGCGGCTGTGGTGGCGCGCGCGCGCGCGCG 600
DB 21246 GGTGTTGACAGAGTGTGTTCCAAAAGAAAAGTAAATCATGATGAGTGGTGTGCGGT 21187
OY 601 TACACGACGCGCGCATGCGCCAAAGGCGATGGCGCGCGCGCTTACGATGATCAAC 660
DB 21186 ACACATGCTGCGCGCATGCGCTTGGTGTGCTCAAGTACTATTTATATATAGT 21127
OY 661 ATCGACAAATTCGGCACTGCAAGCGCGGATTCGTGGCGCGGATCCACACTGCTACTCA 720
DB 21126 TCCACGCTCTCTCACTTCTAGAGAAAGCTTTGGAAGTC AATTCAAACTCTTATGCT 21067
OY 721 TCGGCTACGAGCTCGAGGAGTGGCGTCAAAAGCTCGCGACTGTGATGGCGCTGCTG 780
DB 21066 AATTCATTCAAATGATGACAAAGTGAAGATGCTGATGCTGATGAGCAATTCCTC 21007
OY 781 GTGCCAGGCGCGCAAGCGCAATAGTCTGCAATTCATGCTGGCGGATTTGAAACCA 840
DB 21006 ATCCCTGGTGGCAAAAGCCCGGATTTGGTGAAGATGATGATGATGATGATGATGAT 20947
OY 841 GTGCGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 897
DB 20946 GGTGCTGTA-TGCTTACAGTGTGCTGTTGACAAAGTGGGTTTAAACAGGTGACCT 20888
OY 898 CCGACCACTACGACCAAGCGAGTGTGCGGCTGACAGACAGTGTGTTTACGCTGGGG 957
DB 20887 GTACACAGCAGCATGATACCGCTCTATGAAAACAGGCTTTCACATGCGCTGGCC 20828
OY 958 AACATGCGCGCTCGGCTGGCGGAGAGTGTGATGATGATGATGATGATGATGATGAT 1017
DB 20827 AATATCCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20768
OY 1018 TATGCTGCTGAGCTTGGCGACATGCT 1045
DB 20767 TATATGAAAGCTTTGGCTGCGCAAGAT 20740

RESULT 7
V74442/c
ID V74442 standard; DNA; 9280 BP.
AC V74442;
DE 16-MAR-1999 (first entry)
KW Computer readable medium; vaccine; S. aureus infection; immunodetection;
cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
skin infection; surgical wound infection; scalded skin syndrome;
toxic shock syndrome; ds.
KW skin infection; surgical wound infection; scalded skin syndrome;
toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT FT 841..900
FT FT misc_feature
FT FT /tag-a
FT FT /note- "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FT FT misc_feature
FT FT 2641..2700
FT FT /tag-b
FT FT /note- "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"

```

```

FT misc_feature 4441..4500
FT /tag-c
FT /note- "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FT misc_feature 6241..6300
FT /tag-d
FT /note- "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FT misc_feature 8041..8100
FT /tag-e
FT /note- "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FT EP-786519-A2.
FT 30-JUL-1997.
FT 07-JAN-1997: 100117.
FT 05-JAN-1996: US-009861.
FT (HMA-) HUMAN GENOME SCI INC.
FT Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
FT Rosen CA;
FT WPI: 97-374922/35.
FT Polynucleotide(s) and proteins derived from Staphylococcus aureus
FT stored on computer readable medium and used in the production of
FT anti-S.aureus vaccines
FT Claim 1: Page 710-715: 3271pp. English.
FT This sequence represents one of 5191 Staphylococcus aureus DNA sequences
FT of the invention. The DNA sequences are recorded on a computer readable
FT medium, preferably selected from a floppy or hard disk, random access
FT memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
FT the S.aureus DNA sequences allows putative functions to be assigned so
FT that protein-encoding or regulatory regions of commercial, therapeutic or
FT industrial importance can be obtained. Specifically, sequences which are
FT likely to encode antigens have been identified and these polypeptides can
FT be used in a vaccine composition against S.aureus infection. The
FT polypeptides can also be used in a kit for the immunodetection of
FT S.aureus in a sample. S.aureus is implicated in numerous human diseases,
FT including cellulitis, eyelid infections, food poisoning, osteomyelitis,
FT skin and surgical wound infections, scalded skin syndrome, toxic shock
FT syndrome, etc. Organisms transformed with the DNA sequences can be used
FT for recombinant production of the polypeptides. The new DNA sequences
FT (and their fragments) are useful as primers or probes for isolating
FT homologues of any of the S.aureus DNA sequences contained on the
FT computer readable medium.
FT Sequence 9280 BP; 2958 A; 1649 C; 1309 G; 3060 T;

Query Match 14.5%; Score 178.4; DB 1; Length 9280;
Best Local Similarity 50.6%; Pred. No. 5.5e-29;
Matches 484; Conservative 0; Mismatches 466; Indels 6; Gaps 2;

OY 151 CATGAGTCTATCATCAGACGATCCGAGAGGCTCGGCTATCAGCAGCGGATTTC 210
DB 9280 CTTACTGTTTGTGGAACAAATCGGGTTCAGATCATCTTTGAGAGTATGATTAC 9221
OY 211 AAGCGCGAGCGCGCACTGTGTGGCAACCGCGGACAGGTGTGGCGCGGATTTA 270
DB 9220 AAGAGAGAGGTGCTGAGTGTGTGCTGACAAACAAAGTTTG--GATGTGATATG 9164
OY 271 TTGTCAGAGTAAAGAACCGATAGCGGCGAATAGCGCGCTGCGAGCGAGAGATC 330
DB 9163 GTTATTAAGTTAAAGAACCGATTAATCTGATATTCATATTAAAGAGGGCTTGA 9104
OY 331 TTGTCAGCTTCTTGCATTTGGCGCGTACGCTGCTGACCGATGCTGTTGATTC 390
DB 9103 TTTATCTACTTACTTCTTATGCAATAGAGAAAATTAACCAAGCTTGTATGATGA 9044
OY 391 GGCACCGTCAATTGCTTTCAGAGACGTTCCAGACGCGCGGCGGCACTACCTGCTT 450

```



OY 970 TCGGTCGCGAGAGCTGACCTACGCGCTGACCAACGCGACGATCCGTATGCTCGAG 1029  
 DB 853 ATGTGGCGTAAACGAGCTACGCTTATAGCATGACGAGTGGCGGATTTGTTGAT 912  
 OY 1030 CTGGCCGACCATGCTGCGCGG 1051  
 DB 913 TATTAGAGCATGCTTGAAG 934

## RESULT 9

VT3039/c  
 ID V43039 standard; DNA: 544 BP.  
 AC V43039;  
 DT 09-NOV-1998 (first entry)  
 DE Streptococcus pneumoniae polypeptide coding region.  
 KM Polypeptide; ORF: open reading frame; infection; bacterial;  
 OS Streptococcus pneumoniae; diagnosis; prophylaxis; ds.  
 FH Key Location/Qualifiers  
 FT CDS complement (59..334)  
 FT /\*tag- a  
 FT /note- "polypeptide"  
 MO823631-A1.  
 PN 04-JUN-1998.  
 PD 24-NOV-1997: U21976.  
 PR 21-NOV-1996: US-031879.  
 PA (SMIR ) SMITHKLINE BEECHAM CORP.  
 PI (SMIR ) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Hodgson JE, Knowles DCC, Lonetto MA, Nicholas RO.  
 PI Reid RH, Zarfos PN;  
 PI WPI: 98-322654/28.  
 DR P-PSDB: W62769.  
 DT Streptococcus pneumoniae polynucleotides - useful for developing  
 PT products for diagnosis, prevention and treatment of infections e.g.  
 PT pneumonia, bacteremia, meningitis or endocarditis  
 PS Claim 1: Page 165-166; 18pp; English.  
 CC The sequence is that of a Streptococcal polypeptide coding region.  
 CC The polypeptide can potentially be used for the diagnosis and  
 CC prevention of bacterial infections, especially sp infection.  
 CC It may be used for the treatment of diseases such as otitis media,  
 CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural  
 CC empyema, endocarditis or infection of the cerebrospinal fluid.  
 SQ Sequence 544 BP; 149 A; 121 C; 124 G; 150 T;

Query Match 5.6%; Score 69.2; DB 1; Length 544;  
 Best Local Similarity 53.4%; Pred No 2.4e-06;  
 Matches 191; Conservative 0; Mismatches 163; Indels 4; Gaps 2;

OY 691 TTCTGCGCGCGATCCACACTCGCTACTCATCGGCTACGAGCTCGAGGTCGCTCAA 750  
 DB 543 TTGGAGATCAAAATCAATCTTATATCAATTCATTCACATGGAAGCAATATGAGA 484  
 OY 751 CGTGGCAGACCTGATGAGGCGCTGCTGGCCAGGCGCAGCCCAATATGTC 810  
 DB 483 GATGCTGATGATGATGATGAGCCATCTCATCTCTGTGCAAAAGCACCCTAATGTTG 424  
 OY 811 TCGAATTAATCTGTCGCGCATATGAACAGGTCGCTACTGCTGATATAGCATGAC 870  
 DB 423 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 365  
 OY 871 CAGGCGGCTGTTTGA--AGGCTCAGACGACGACAC TACGACACACCGACGTCGCC 927  
 DB 364 CAAAGTGGGCTTATGAAACAGCTGATGACAGCAGCAGATGAACCGTCTATGAA 305  
 OY 928 GTGCAAGACAGCTGTTTACTGCTGGGGAACATGCCGCTCGGTCGAGACGTCG 987  
 DB 304 AAGACAGGCTTCTCCACATGCGCTGCGCATATCTCTGCGGTTGCCGCACTCA 245  
 OY 988 ACCTACGCGCTGACCAACGCGATGCGGATGCTGCTGACGCTTCCGACCATGCT 1045  
 DB 244 ACCATGCGCCTAAGCAATGTCACCTCTCTATATGGAAGCTTGGCGGCAAGAT 187

## RESULT 10

VT3039/c  
 ID V74370 standard; DNA: 31096 BP.  
 AC V74370;  
 DT 16-MAR-1999 (first entry)  
 DE Streptococcus aureus contig SEQ ID #59.  
 KM Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KM skin infection; surgical wound infection; scalded skin syndrome;  
 OS Streptococcus aureus.  
 FH Key Location/Qualifiers  
 FT misc-feature 1201..1260  
 FT /\*tag- a  
 FT /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"  
 misc-feature 3001..3060  
 FT /\*tag- b  
 FT /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"  
 misc-feature 4801..4860  
 FT /\*tag- c  
 FT /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"  
 misc-feature 6601..6660  
 FT /\*tag- d  
 FT /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"  
 misc-feature 8401..8460  
 FT /\*tag- e  
 FT /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"  
 misc-feature 10201..10260  
 FT /\*tag- f  
 FT /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"  
 misc-feature 12001..12060  
 FT /\*tag- g  
 FT /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"  
 misc-feature 13801..13860  
 FT /\*tag- h  
 FT /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"  
 misc-feature 15601..15660  
 FT /\*tag- i  
 FT /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"  
 misc-feature 17401..17460  
 FT /\*tag- j  
 FT /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"



FT misc-feature 19201. 19260  
 FT /tag- k  
 FT "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT misc-feature 21001. 21060  
 FT /tag- l  
 FT "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT misc-feature 22801. 22860  
 FT /tag- m  
 FT "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT misc-feature 24601. 24660  
 FT /tag- n  
 FT "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT misc-feature 26401. 26460  
 FT /tag- o  
 FT "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT misc-feature 28201. 28260  
 FT /tag- p  
 FT "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT misc-feature 30001. 30060  
 FT /tag- q  
 FT "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT misc-feature 30001. 30060  
 FT /tag- g  
 FT "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT BP-786519-A2.  
 FT 30-JUL-1997.  
 FT 07-JAN-1997: 100117.  
 FT 05-JAN-1996: US-009861.  
 FT (HUMAN-) HUMAN GENOME SCI INC.  
 FT Barash SC, Choi GH, Dillon PJ, Fannon MR, Kuncsch CA,  
 FT Rosen CA.  
 FT WPI; 97-374922/35.  
 FT Polynucleotide(s) and proteins derived from *Staphylococcus aureus*  
 FT stored on computer readable medium and used in the production of  
 FT anti-S aureus vaccines  
 FT Claim 1; Page 452:469; 321pp; English  
 FT This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences  
 FT of the invention. The DNA sequences are recorded on a computer readable  
 FT medium, preferably selected from a floppy or hard disk, random access  
 FT memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 FT the S aureus DNA sequences allows putative functions to be assigned so  
 FT that protein-encoding or regulatory regions of commercial, therapeutic or  
 FT industrial importance can be obtained. Specifically, sequences which are  
 FT likely to encode antigens have been identified and these polypeptides can  
 FT be used in a vaccine composition against S aureus infection. The  
 FT polypeptides can also be used in a kit for the immunodetection of  
 FT S aureus in a sample. S aureus is implicated in numerous human diseases,  
 FT including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 FT skin and surgical wound infections, scalded skin syndrome, toxic shock  
 FT syndrome, etc. Organisms transformed with the DNA sequences can be used  
 FT for recombinant production of the polypeptides. The new DNA sequences  
 FT (and their fragments) are useful as primers or probes for isolating  
 FT CC homologues of any of the S aureus DNA sequences contained on the  
 FT CC computer readable medium.  
 FT Sequence 31096 BP; 11857 A; 5243 C; 5477 G; 7488 T;

Query Match 4.7%: Score 57.6; DB 1: Length 31096;  
 Best Local Similarity 44.9%: Pred. No. 0.00084;  
 Matches 440; Conservative 0; Mismatches 519; Indels 21; Gaps 5;  
 QY 78 GACGAGACCAAAACCAAGATTCGGGTGGCCCTACCCCGCGCGCTCGGGAAC 137  
 DB 10 GCAAGATTTAAACAGAGTGAAGCGCTGTAGCTTGACACCCGAAATGTCGTA 69  
 QY 138 AACCGGTGGCCATGAGGTGCTATCCAGGACAGTCCCGAGAGGCTCGCTATC 197  
 DB 70 AACGATCGAGACATAAAGTATGTTGAAAAAATGCTGGCATTTGGTCAAGATTTC 129  
 QY 198 CGACCGGATTCACAGGCTCGGCGCCGCACTGGTGGGACCCGCGACGAGTGGGC 257  
 DB 130 TAAAGTATGTGTAAGAAAGAGGCGCTAA--GATCGTACTACGAAACATG-- 185  
 QY 258 CGACGCTGATTTATGCTCAAGGTCAAGAACCGATAGCGGGAATAGCGCGCTGCG 317  
 DB 185 -GAGGTGATCTGTGTTATCAAGTAAGTAAGAACCTCATGAAAGCAATATTTCAA 243  
 QY 318 ACAGGCGCATCTGTTCTTCTGCTTTGGCCGCGTACGCTGCTGACCGATGC 377  
 DB 244 AAGAAATCAAAATTAATCTGGGATTTTACATCTACATCTTCAAAAGAAATAGTAA 303  
 QY 378 GTTGTGATTCGGGACACGCTCAATTGCTTACGAGACCGTCCAGACCGCGCGCGC 437  
 DB 304 AATGCAAGAGTGTGTTGTTAGTACGCTGTTAGTGTGTAACCAAT--TATTAAGAA 360  
 QY 438 ACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 497  
 DB 361 AGCAGATATTAAGCGCCCAATGAGTGTATAGCAGTCAACCTCAGCAATATTTGG 420  
 QY 498 TTACACCTGATGCAAGCAACCAAGGCGCGGCTGCTGATGAGGCGGCGCGCGG 557  
 DB 421 TTACTACTGTGAAGCAACAATGAGTGTGTAAGTACTTATGACTGGTATCATGAAA 480  
 QY 558 CGAACCGCGCGACG-----TCGTGTGATGCGCGCGCGCGCGCGCGCTACMAC 608  
 DB 481 TGTGATATACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 540  
 QY 609 AGCCGCGATGCGCAAGCGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 668  
 DB 541 AGCAAAATGTTGCTTGGGACTAAATGCTAAATGCTAAATGCTAAATGCTAAAT 600  
 QY 669 ACTGCGCACTGACGCGCGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 725  
 DB 601 CATTAATATCTTGAAGATATGATGATGATGATGATGATGATGATGATGATGAT 660  
 QY 726 CTACGAGCTGAGGAGTGCCTCAAGCGTGCAGCTGTGATTTGGGCGGTCTGTG 785  
 DB 661 AGAAATTTAGCAGACCAATTAAGAAACAGATGATTTATTTTCAATTTTATTTTC 720  
 QY 786 AGCGCGCAAGCGCCCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 845  
 DB 721 AGTGGGAAACCGCCCAATTTGTTACTGTGAGATGTTAATTAATTAATTAATTA 780  
 QY 846 GTTACGTTGATATAGCATGACCGAGCGGCGGCTTTTCGAAGGCTCGACCGCAC 905  
 DB 781 AGTATTAATGATATAGCATGATTAATTAATTAATTAATTAATTAATTAATTA 840  
 QY 906 CTACGACACCGCGCGCTTCCGCTGACGACAGCGCTTTTACTGCTGCGCAATG 965  
 DB 841 AATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900  
 QY 966 CGCTGCTGCGCAAGAGTGCAGTACGAGCGTGCAGCAACGCGACGATCGCTATG 1025  
 DB 901 AGAGGAGTGCAGCAAGCACTCAACCAATGCGATTAAGCAAGAAATATGATTA 960  
 QY 1026 CGAGCTTCCGCGCAATGCT 1045  
 DB 961 AGAAATTTGTGACAAAGGCT 980







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:07:29 ; Search time 5541.94 Seconds  
(without alignments)  
898.126 Million cell updates/sec

Title: US-09-362-485-5  
Sequence: 1228  
1 ATCTTCGAGATTAATCGAAC.....GGGAGTAGGGAAGCGATGA 1228

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 4857316 seqs, 2026611550 residues  
Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

- 1: em\_est1:\*
- 2: em\_est2:\*
- 3: em\_est3:\*
- 4: em\_est4:\*
- 5: em\_est5:\*
- 6: em\_est6:\*
- 7: em\_est7:\*
- 8: em\_est8:\*
- 9: em\_est9:\*
- 10: em\_est10:\*
- 11: em\_est11:\*
- 12: em\_est12:\*
- 13: em\_est13:\*
- 14: em\_est14:\*
- 15: em\_est15:\*
- 16: em\_est16:\*
- 17: em\_est17:\*
- 18: em\_est18:\*
- 19: em\_est19:\*
- 20: em\_est20:\*
- 21: em\_est21:\*
- 22: em\_est22:\*
- 23: em\_est23:\*
- 24: em\_est24:\*
- 25: em\_est25:\*
- 26: em\_est26:\*
- 27: em\_est27:\*
- 28: em\_est28:\*
- 29: em\_est29:\*
- 30: em\_est30:\*
- 31: em\_est31:\*
- 32: em\_est32:\*
- 33: em\_est33:\*
- 34: em\_est34:\*
- 35: em\_est35:\*
- 36: em\_est36:\*
- 37: em\_est37:\*
- 38: em\_est38:\*
- 39: em\_est39:\*
- 40: em\_est40:\*
- 41: em\_est41:\*
- 42: em\_est42:\*
- 43: em\_est43:\*
- 44: em\_est44:\*
- 45: em\_est45:\*

- 45: gb\_est26:\*
- 46: gb\_est27:\*
- 47: gb\_est28:\*
- 48: gb\_est29:\*
- 49: gb\_est30:\*
- 50: gb\_est31:\*
- 51: gb\_est32:\*
- 52: gb\_est33:\*
- 53: gb\_est34:\*
- 54: gb\_est35:\*
- 55: gb\_est36:\*
- 56: gb\_est37:\*
- 57: gb\_est38:\*
- 58: gb\_est39:\*
- 59: gb\_est40:\*
- 60: gb\_est41:\*
- 61: gb\_est42:\*
- 62: gb\_est43:\*
- 63: gb\_est44:\*
- 64: gb\_est45:\*
- 65: gb\_est46:\*
- 66: gb\_est47:\*
- 67: gb\_est48:\*
- 68: gb\_est49:\*
- 69: gb\_est50:\*
- 70: gb\_est51:\*
- 71: gb\_est52:\*
- 72: gb\_est53:\*
- 73: gb\_est54:\*
- 74: gb\_est55:\*
- 75: gb\_est56:\*
- 76: gb\_est57:\*
- 77: gb\_est58:\*
- 78: gb\_est59:\*
- 79: gb\_est60:\*
- 80: gb\_est61:\*
- 81: gb\_est62:\*
- 82: gb\_est63:\*
- 83: gb\_est64:\*
- 84: gb\_est65:\*
- 85: gb\_est66:\*
- 86: gb\_est67:\*
- 87: gb\_est68:\*
- 88: gb\_est69:\*
- 89: gb\_est70:\*
- 90: gb\_est71:\*
- 91: gb\_est72:\*
- 92: gb\_est73:\*
- 93: gb\_est74:\*
- 94: gb\_est75:\*
- 95: gb\_est76:\*
- 96: gb\_est77:\*
- 97: gb\_est78:\*
- 98: gb\_est79:\*
- 99: gb\_est80:\*
- 100: gb\_est81:\*
- 101: gb\_est82:\*
- 102: gb\_est83:\*
- 103: gb\_est84:\*
- 104: gb\_est85:\*
- 105: gb\_est86:\*
- 106: gb\_est87:\*
- 107: gb\_est88:\*
- 108: gb\_est89:\*
- 109: gb\_est90:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

[illegible]







FEATURES	location/Qualifiers
Source	1. 846
	/organism="Drosophila melanogaster"
	/plasmid="pBelobAC11"
	/db_xref="taxon:7227"
	/clone_lib="DrosBAC"
	/clone="BACN04N13"
	/note="end : 17"
BASE COUNT	300 a 7 c 40 g 175 t 284 others
ORIGIN	
Query Match	3.8%; Score 46.6; DB 82; Length 846;
Best Local Similarity	13.0%; Pred. No. 6.7;
Matches	33; Conservative 127; Mismatches 93; Indels 0; Gaps 0;
OY	363 TCGTTCACCGATGCGCTGTTGGATTCCGGACACACGTCATATGCTTACGACGACGCTCCA 422
DB	842 TACTCTSSSTSSCTSSSSSTBTBTSTSTASSSSSSTSSSTSSSTSSSTSSSTBTBT 783
OY	423 GACCCCGGACGCGACATCCCTGCTGCTCCCGATGACGACGATGCGCGGCGACATCGC 482
DB	782 SSSSTTTTSSSTSSSTTSSSTTSSSTGCTSSCSCSSSSSTTS:SASASSTTSSSTTG:SBS 723
OY	483 CGCCGAGTTGGCGGCTTACCACTGATGCGAACCCAGAGGGGCGCGGCTGCTGATGGG 542
DB	722 SBTGTSSTSBTBTBTBTSSBSBSRSSSSSTSTBTBTSSBSBSSTSSSTSS 663
OY	543 CGGGGCGCCGCGCGTACGCGGCGGACGCGGCGGATCGGCGGCGCGGCGCGGCTTA 602
DB	662 SSSSBTSSSTSSSTSTCTTSTTSSCSCSSSSSTSSSTSSBSBSBSSTSSSCT 603
OY	603 CAACGCGACCCGC 615
DB	602 CGCGSCGCCCC 590
RESULT 7	
CNS0181E/c	
LOCUS	CNS0181E 1101 bp DNA GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37H05 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL108764
VERSION	AL108764.1 GI:5629068
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 1101)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) + http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CERH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.
FEATURES	Location/Qualifiers
Source	1. 1101
	/organism="Drosophila melanogaster"
	/plasmid="pBelobAC11"
	/db_xref="taxon:7227"
	/clone_lib="DrosBAC"
	/clone="BACN37H05"

[illegible]

```

RESULT      B
LOCUS       CINS006XK      935 bp      DNA      GSS      03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACRLAN09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066051
VERSION     AL066051.1  GI:4945019
KEYWORDS
SOURCE
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 935)
            Genoscope.
            Direct Submission
            Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENTARY
            - Web : www.genoscope.cns.fr)
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mamosser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain Y2, cn bw sp; the same strain used for the bdgp's
            p1 and EST libraries. A more detailed description of the library,
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES
            source
            1..935
            Location/Qualifiers
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone_lib="RPCI-98"
            /clone="BACRLAN09"
            /note="end : 77"
BASE COUNT  257 a      170 c      162 g      96 t      250 others

```





d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

FEATURES  
source  
Location/Qualifiers  
1..744  
/organism="Drosophila melanogaster"  
/plasmid="pBelOBAC11"  
/db\_xref="taxon:7227"  
/clone\_lib="DrosBAC"  
/clone="BACN37B10"  
/note="end : SP6"

BASE COUNT 20 a 191 c 245 g 105 t 183 others  
ORIGIN

Query Match 3.7%; Score 45.2; DB 83; Length 744;

Best Local Similarity 34.0%; Pred. No. 12;  
Matches 87; Conservative 41; Mismatches 128; Indels 0; Gaps 0;

OY 329 TCTTGTACGCTTCTTGCATTTGGCCGCGTCACGCTGCTTGCACGATGCTGTGAT 388  
Db 191 YCTTGTCTCTTCTTGCATTTGGCCGCGTCACGCTGCTTGCACGATGCTGTGAT 250  
OY 389 CCGGACACGCTCAATTCCTACGACGCTCAGACGCTCAGACGCTCAGACGCTCAG 448  
Db 251 YCTTGTCTCTTCTTGCATTTGGCCGCGTCACGCTGCTTGCACGATGCTGTGAT 310  
OY 449 TTGCCCCGATGACGAGTCCGCGTGCATGCGCCGCGCAGGTTGAGCTTACCACTGA 508  
Db 311 CCGCCCCG 370  
OY 509 TCGGAACCCAGGGGGGGGGTGTGATGAGGGGGGGTGTGATGAGGGGGGGTGTG 568  
Db 371 CCGGACG 430  
OY 569 ACGTCTGTGTATCGG 584  
Db 431 NGGNGSGSGCTKTGG 446

## RESULT 13

CNS0175Y 1101 bp DNA GSS 26-JUL-1999  
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC  
DEFINITION BACN37B108 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL108460 GI:5628764  
VERSION AL108460.1 GI:5628764  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 Evry cedex - FRANCE (E-mail : sequefgenoscope.cns.fr  
COMMENT Web : www.genoscope.cns.fr  
determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelOBAC11.

FEATURES  
source  
Location/Qualifiers  
1..1101  
/organism="Drosophila melanogaster"  
/plasmid="pBelOBAC11"

/db\_xref="taxon:7227"  
/clone\_lib="DrosBAC"  
/clone="BACN37B108"  
/note="end : SP6"

BASE COUNT 254 a 176 c 160 g 152 t 359 others  
ORIGIN

Query Match 3.7%; Score 45.2; DB 83; Length 1101;

Best Local Similarity 12.6%; Pred. No. 13;  
Matches 40; Conservative 157; Mismatches 121; Indels 0; Gaps 0;

OY 420 CCAGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 479  
Db 755 CGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 814  
OY 480 CGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 539  
Db 815 CSVASASASVMSKVASAVASGASVAVASGASVAVASGASVAVASGASVAVAS 874  
OY 540 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 599  
Db 875 SSSSVSMAAASVSSASASASMAAASVAVASVAVASVAVASVAVASVAVAS 934  
OY 600 CTACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 659  
Db 935 SVASASVSSSSSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSS 994  
OY 660 CATGACAACTTGCGCACTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 719  
Db 995 AAAAAAASASVAVASVAVASVAVASVAVASVAVASVAVASVAVASVAVAS 1054  
OY 720 ATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 737  
Db 1055 VSAVAV SASASVSSVA 1072

## RESULT 14

A1712257 315 bp mRNA EST 02-FEB-2000  
LOCUS 605069D09.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays  
DEFINITION cDNA, mRNA sequence.  
ACCESSION A1712257  
VERSION A1712257.1 GI:5065195  
KEYWORDS EST.  
SOURCE Zea mays.  
ORGANISM Zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Zea.  
REFERENCE 1 (bases 1 to 315)  
AUTHORS Walbot V.  
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford  
JOURNAL University  
COMMENT Unpublished (1999)  
On Jun 22, 1998 this sequence version replaced gi:3246668.  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 605069 row: D column: 09.

FEATURES  
source  
Location/Qualifiers  
1..315  
/organism="Zea mays"  
/cultivar="Ohio43"  
/db\_xref="taxon:4577"  
/clone\_lib="605 - Endosperm cDNA library from Schmidt lab"  
/tissue\_type="nucellar, embryo, and endosperm"  
/dev\_stage="10-14 days post-pollination"  
/lab\_host="DH5(alpha)"





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:07:29 ; Search time 5541.94 Seconds  
(without alignments)  
898.126 Million cell updates/sec

Title: US-09-362-485-5  
Perfect score: 1228  
Sequence: 1 ATCTGCAGATTAATCGAAC.....GGGAGTAAGGAGGAGCATGA 1228

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 202661650 residues  
Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

ESR:\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: gb\_est33:\*  
53: gb\_est34:\*  
54: gb\_est35:\*  
55: gb\_est36:\*  
56: gb\_est37:\*  
57: gb\_est38:\*  
58: gb\_est39:\*  
59: gb\_est40:\*  
60: gb\_est41:\*  
61: gb\_est42:\*  
62: gb\_est43:\*  
63: gb\_est44:\*  
64: gb\_est45:\*  
65: gb\_est46:\*  
66: gb\_est47:\*  
67: gb\_est48:\*  
68: gb\_est49:\*  
69: gb\_est50:\*  
70: gb\_est51:\*  
71: gb\_est52:\*  
72: gb\_est53:\*  
73: gb\_est54:\*  
74: gb\_est55:\*  
75: gb\_est56:\*  
76: gb\_est57:\*  
77: gb\_est58:\*  
78: gb\_est59:\*  
79: gb\_est60:\*  
80: gb\_est61:\*  
81: gb\_est62:\*  
82: gb\_est63:\*  
83: gb\_est64:\*  
84: gb\_est65:\*  
85: gb\_est66:\*  
86: gb\_est67:\*  
87: gb\_est68:\*  
88: gb\_est69:\*  
89: gb\_est70:\*  
90: gb\_est71:\*  
91: gb\_est72:\*  
92: gb\_est73:\*  
93: gb\_est74:\*  
94: gb\_est75:\*  
95: gb\_est76:\*  
96: gb\_est77:\*  
97: gb\_est78:\*  
98: gb\_est79:\*  
99: gb\_est80:\*  
100: gb\_est81:\*  
101: gb\_est82:\*  
102: gb\_est83:\*  
103: gb\_est84:\*  
104: gb\_est85:\*  
105: gb\_est86:\*  
106: gb\_est87:\*  
107: gb\_est88:\*  
108: gb\_est89:\*  
109: gb\_est90:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES

No.	Score	Match	Length	DB	ID	Description
1	57.4	4.7	925	82	CNS0091P	AL053013 Drosophila
2	55.2	4.5	925	82	CNS0091P	AL053013 Drosophila
3	51.8	4.2	535	74	AV390505	AV390505 AV390505
4	50.2	4.1	844	82	CNS0052P	AL056652 Drosophila
5	49.0	4.0	910	82	CNS0060N	AL056629 Drosophila
6	46.6	3.8	846	82	CNS010R	AL099337 Drosophila
7	46.6	3.8	1101	82	CNS0181E	AL108764 Drosophila
8	46.4	3.8	935	82	CNS006XK	AL105599 Drosophila
9	46.2	3.8	645	82	CNS01213	AL066051 Drosophila
10	45.8	3.7	935	82	CNS006XK	AL108764 Drosophila
11	45.4	3.7	591	69	AM128786	AM128786 f37/d05.y
12	45.2	3.7	744	83	CNS0172K	AL108764 Drosophila
13	45.2	3.7	1101	83	CNS0172X	AL108764 Drosophila
14	45.2	3.7	1101	83	CNS0172X	AL108764 Drosophila
15	45.2	3.7	1101	83	CNS0172X	AL108764 Drosophila
16	44.8	3.6	846	82	CNS010R	AL099337 Drosophila
17	44.4	3.6	932	82	CNS00720	AL059666 Drosophila
18	43.8	3.6	843	82	CNS006XK	AL108764 Drosophila
19	43.8	3.6	427	51	AI736678	AI736678 sb32b04.y
20	43.8	3.6	512	74	AV396794	AV396794 AV396794
21	43.8	3.6	521	74	AV397116	AV397116 AV397116
22	43.6	3.6	526	74	AV394827	AV394827 AV394827
23	43.2	3.5	1101	83	CNS010R	AL099337 Drosophila
24	43.2	3.5	439	81	AO864063	AO864063 nbe0022M
25	43.2	3.5	798	82	CNS006XK	AL055851 Drosophila
26	41.8	3.4	414	80	AM286084	AM286084 LG1_261.B
27	41.8	3.4	465	80	AM287007	AM287007 LG1_264.C
28	41.8	3.4	530	74	AV396353	AV396353 AV396353
29	41.8	3.4	839	82	CNS004NB	AL066742 Drosophila
30	41.8	3.4	932	82	CNS00720	AL059666 Drosophila
31	41.6	3.4	506	74	AM180713	AM180713 MGA0863f
32	41.6	3.4	1101	82	CNS006XK	AL064271 Drosophila
33	41.4	3.4	1101	82	CNS006XK	AL064271 Drosophila
34	41.4	3.4	1201	83	CNS014B1	AL103945 Drosophila
35	41.2	3.4	639	60	AI812194	AI812194 605086F10
36	41.2	3.4	774	105	AO327321	AO327321 nbe0004OD
37	41.2	3.4	910	82	CNS0060N	AL056629 Drosophila
38	41.1	3.3	605	46	AI461529	AI461529 466018G05
39	41.1	3.3	1101	83	CNS0175X	AL108460 Drosophila
40	40.8	3.3	384	64	AM054773	AM054773 w559d08.x
41	40.8	3.3	385	21	D48746	D48746 R1C15137A
42	40.8	3.3	437	74	AV395790	AV395790 AV395790
43	40.6	3.3	506	74	AV396405	AV396405 AV396405
44	40.4	3.3	574	91	AO848186	AO848186 LMAJFV1.1
45	40.2	3.3	916	83	CNS01550	AL105860 Drosophila

## ALIGNMENTS

RESULT 1  
CNS0091P/c  
LOCUS  
DEFINITION  
CNS0091P 925 bp DNA  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AL053013  
AL053013.1 GI:4934461  
GSS  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
1 (bases 1 to 925)

AUTHORS  
TITLE  
JOURNAL  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

COMMENT  
Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Ooeegawa and  
Aron Mammone at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
restriction digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain Y2: cn bw sp, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source  
1..925  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACR19D16"  
/note="end : TET3"

BASE COUNT  
120 a 61 c 61 g 172 t 511 others

ORIGIN

Query Match 4.7% Score 57.4; DB 82; Length 925;

Best Local Similarity 11.9%; Pred. No. 0.055; Mismatches 142; Indels 0; Gaps 0;

Matches 43; Conservative 176; Mismatches 142; Indels 0; Gaps 0;

442 CCCCTCTGCGCCGATGAGCAGATCGCGGTCGACCTGCCCGCCAGGTTGGCCTTAC 501

920 SCSCSCSBBSCSSSSSMTSSNSBSCSSBSSSTSSMSSSSSSSSSSSSSSSSS 861

502 CACCTGATGCGAACCAGAGGGGCGCGTGTGCTGATGGCGGGGTGCCCGCTGAA 561

860 SACKVNMASSCCGCCGABCCMCSSSSCCGSASARVKKVASGAGKRGGGGASA 801

562 CCGCGACGATCGTGTGATCGCGCGGACCGCGGCTACATCGACCGCATCCGC 621

800 SHSSSSACBSSSSSSCSAS...SASSSSASSSSRRSRSGGAGGSR, SSRSSSSSSASAGS 741

622 AAGCGATGGCGCGGACCGTACGCTTACATCAGATCAGACAACTTGGGAATCC 681

740 :SS 681

682 GAGCGGATGCTTGGCGCGGATCCACTCGTACTCATCGGCTTACAGAGGCT 741

680 SCSTASMSAARSS 621

742 GCCGTCAACGTCGCGACCTGATGAGGCGCGCTTGTGCGGACGCGCCAGGACCC 801

620 GGGSSVASSGSS 561

802 A 802

560 R 560

RESULT 2

CNS0091P  
LOCUS  
DEFINITION  
CNS0091P 925 bp DNA  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AL053013  
AL053013.1 GI:4934461  
GSS  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
1 (bases 1 to 925)





FEATURES	pbeloBAC11.	
SOURCE	1. 846	
	/organism="Drosophila melanogaster"	
	/plasmid="pbeloBAC11"	
	/db_xref="taxon:7227"	
	/clone_1lb="DrosBAC"	
	/clone="BACN04N13"	
	/note="end : T7"	
BASE COUNT	300 a	7 c 40 g 175 t 284 others
ORIGIN		
Query Match	3.8%.	Score 46.6; DB 82; Length 846;
Best Local Similarity	13.0%.	Pred. No. 6.7;
Matches	33; Conservative	127; Mismatches 93; Indels 0; Gaps 0;
QY	363	TGCTTGACCGATGCGTTGTGGATTCCGCACACATGTCATGCTACGAGACCGTCCA 422
Db	842	TTCTSSSTSSSTSCSSSSSTBTSTTSTSTSSASSSSSSSTSSSTSSSTSSSTSSBT 783
QY	423	GACCCCGACGGCGGACACACACCCCTGCGCCCGGATGAGACAGTCGGCGGCGATGC 482
Db	762	SBSSTSTSTSSSTSTSTSSSTSTSCSSSSSTSSSASASSTSTSTTCSSBS 723
QY	483	CGCCGAGGTGCGCTTACACACCTGATCGGACCCAGAGGGCGCGGTGTCGTATGG 542
Db	722	SBSTGTSSTSBTBTBTBTSSSSBSBSSTSSSTBTBTSSSSBSSTSTSS 663
QY	543	CGGGGTGCCGGCGTCGAAACCGCGGACGTCGTGATCGGCGCGGACCGCGGCTA 602
Db	662	SSSBSTSTSSBSSTCTTSTTSSSCSSSTSTSSBSBSSTSSBSSTSSCST 603
QY	603	CAACGACGCGCCG 615
Db	602	CGCCGCGCCCGCC 590
RESULT 7		
CNS0181E/c		
LOCUS	CNS0181E	1101 bp DNA GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC	
	BACN37H05 of DrosBAC library from Drosophila melanogaster (fruit	
	fly), genomic survey sequence.	
ACCESSION	AL108764	
VERSION	AL108764.1 GI:5629068	
KEYWORDS	GSS.	
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 1 to 1101)	
AUTHORS	Genoscope.	
TITLE	Direct Submission	
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de sequence ;	
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr	
	- Web : www.genoscope.cns.fr)	
COMMENT	Determination of this BAC end sequence was carried out as part of a	
	collaboration with the European Drosophila genome project (EDGP) :	
	http://www.edgp.btl.ac.uk -. This Drosophila melanogaster BAC	
	library (Dros BAC) was made by Alain Billaud at CEPH (Centre	
	d'Etude du Polymorphisme Humain) with funding provided by a MRC	
	project grant. The DNA was prepared from embryos by Alain Bucheton	
	and Genevieve Payan. It has been constructed in the vector	
	pbeloBAC11.	
FEATURES	Location/Qualifiers	
source	1. 1101	
	/organism="Drosophila melanogaster"	
	/plasmid="pbeloBAC11"	
	/db_xref="taxon:7227"	
	/clone_1lb="DrosBAC"	
	/clone="BACN37H05"	

BASE COUNT	61 a	215 c	301 g	235 t	289 others
ORIGIN	/note="end : SP6"				
Query Match	3.8%; Score 46.6; DB 83; Length 1101;				
Best Local Similarity	40.2%; Pred. No. 6.9;				
Matches 101; Conservative	40; Mismatches 109; Indels 1; Caps 1;				
OY	389	CCGGACACAGCTCATTTCCCTACGACAGCCGTCACAGCCGCGGACGGCGCATACCCCGC	448		
Db	308	CCCCCCTTTTMMVMMMAAACCCCAVASGSSCCCCSSCCSSCCGCGSCSACSCSCSSG	249		
OY	449	TTGCCCGCATAGAGCAATCCCGGCTGCACCTCGCCGCGCCAGATTGGCGCTTACCACTGA	508		
Db	248	AAGMGSGMSVGGCCACAAAGGCGGAGACCCGCGCAGCSCCCSSGCGSSCCSSGSSSS	189		
OY	509	TGCGAACCCAAAGGGGGCGCGGTGTGCTGATGAGGGGGGTGCCCGCTCGAACGGCGG	568		
Db	188	CGCGACCCSSSAGGGGCCCGCGCCGCCACCCCGCGSAGCVCAGCCGCAC-CMMVGGSG	130		
OY	569	ACGTGATGTGATCGCGCCCGACCGCCGCGGTCTACACGACGACCCGCAATGCGCAAGGCA	628		
Db	129	ACGCAAGGAGGAGAGAGGGCGCGGAGCCCGCAGCCACCCCGCGCGCCAGCGCGGCGCA	70		
OY	629	TGGGCGCGCAGC	639		
Db	69	ACGCCCAAC	59		
RESULT	8				
CNS006XK	935 bp DNA				
LOCUS	CNS006XK 935 bp DNA GSS 03-JUN-1999				
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC #				
LOCUS	BAC14009 of RPCI-98 library from Drosophila melanogaster (fruit				
DEFINITION	fly), genomic survey sequence.				
ACCESSION	AL065051				
VERSION	AL065051.1 GI:4945019				
KEYWORDS	GSS.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
AUTHORS	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;				
TITLE	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
JOURNAL	1 (bases 1 to 935)				
COMMENT	Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequenage BP 191 91006 EVRY Coudex - FRANCE (E-mail : seq-ref@genoscope.cns.fr - Web : www.genoscope.cns.fr) determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammot in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the 18000 strain y2; cn bw sp, the same strain used for the ADP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .				
FEATURES	Location/Qualifiers				
source	1..935 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPCI-98" /clone="BAC14009" /note="end : T7"				
BASE COUNT	257 a	170 c	162 g	96 t	250 others



## ORIGIN

Query Match 3.8%: Score 46.4; DB 82; Length 935;  
 Best Local Similarity 27.7%: Pred. No. 7.4;  
 Matches 77; Conservative 76; Mismatches 125; Indels 0; Gaps 0;

368 GCACGATGCTTGTGATCCGACACGATGATGCTTACGAGACCGTCCAGACCG 427  
 Db 511 SCMKRTSGSTGCTCCGMMSSSVSCSGCTCCGCTSCMKGCTGCKGCGCG 570  
 Oy 428 CCGAGCGGCGACTTACCTGCTGCTGCGGATGAGCGAAGTGGCCGCTGACTGCGCCCG 487  
 Db 571 CTSCTSSSSCSBSSTCTCTBCTKCSGCTGSGCTGCCGCGGCGCGCGCGCG 630  
 Oy 488 AGCTTGGGGCTTACACGATGATGCGAACCAGGGGGCGCGGCTGCTGATGGCGGG 547  
 Db 631 SGGCGGCGCGSSGSGSSGSSGSSGSSGSGSGGGGGGCGCGCGCGCGCGCG 690  
 Oy 548 TGCCCGGCGTCCGACCGCGCGACGCTGCTGATGCGCGCGCGCGCGCGCGCG 607  
 Db 691 GSCGCGSSGCG 750  
 Oy 608 CAGCCCGCATGCCCAACGCGCATGGCGCGCGCGCGCGCGCGCGCGCGCG 645  
 Db 751 SSGCG 788

## RESULT 9

CNS01213/c 645 bp DNA GSS ? -JUL-1999  
 LOCUS Drosophila melanogaster genome survey sequence T7 d of BAC  
 DEFINITION BACN08007 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL101589  
 VERSION AL101589.1 GI:56132200  
 KEYWORDS GSS.

ORGANISM Drosophila melanogaster  
 SOURCE fruit fly.  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 645)  
 AUTHORS Genoscope.

COMMENT Direct Submission  
 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)  
 - Web : www.genoscope.cns.fr

FEATURES  
 Location/Qualifiers  
 1..645  
 /organism="Drosophila melanogaster"  
 /plasmid="pBelobAC11"  
 /db\_xref="taxon:7227"  
 /clone\_lib="DrosBAC"  
 /clone="BACN08007"  
 /note="end : 17"

BASE COUNT 28 a 26 c 85 g 92 t 414 others  
 ORIGIN

Query Match 3.8%: Score 46.2; DB 82; Length 645;  
 Best Local Similarity 10.3%: Pred. No. 7.7;  
 Matches 43; Conservative 135; Mismatches 238; Indels 3; Gaps 1;  
 Oy 311 GGCTGGAGACGGGAGATCTGTCTGCTGCTGCGCGCGCGCGCTGCA 370

Db 633 SSSSSNN 574

Oy 371 CCATGCTGTTGGATTCGGGCGACGATGATGCTTACGAGACCGTCCAGACCGCG 430

Db 573 CAGSNNCCGNN 514

Oy 431 ACAGCGCATACCTCTGCTGCGGATGAGCGAAGTGGCGGCTGCGCGCGCGCG 490

Db 513 ATNNNNCTANANANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 454

Oy 491 TTGGCGCTTACCATCTGATGCGAACCAGGGGGCGCGGCTGCTGATGAGCGG 550

Db 453 NNNSTSSNNAAAASSSSCASNNSSSSASASASASASASASASASASASAS 394

Oy 551 CCGCGCTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 607

Db 393 SSS 334

Oy 608 CAGCCCGCATGCCCAACGCGCATGGCGCGCGACCGTTCAGCTTACATCAATCGACA 667

Db 333 SSS 274

Oy 668 AACTTGGCAACTGACCGCGATGCTGCGCGCGCGCGCGCGCGCGCGCGCG 726

Db 273 NSSS 215

## RESULT 10

CNS006XK/c 935 bp DNA GSS 03-JUN-1999  
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #  
 DEFINITION BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL066051  
 VERSION AL066051.1 GI:4945019  
 KEYWORDS GSS.

ORGANISM Drosophila melanogaster  
 SOURCE fruit fly.  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 935)  
 AUTHORS Genoscope.

COMMENT Direct Submission  
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)  
 - Web : www.genoscope.cns.fr

FEATURES  
 Location/Qualifiers  
 1..935  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="RPCI-98"  
 /clone="BACR14N09"  
 /note="end : 17"

BASE COUNT 257 a 170 c 162 g 96 t 250 others  
 ORIGIN







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:53:39 ; Search time 8627.09 Seconds  
(with out alignments)  
-139.259 Million cell updates/sec

Title: US-09-362-485-6  
Perfect score: 1235  
Sequence: 1 ATCTGCGAGTATATCGAAC.....GGGAGCGATGATCGGCC 1235

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_bal:\*  
2: gb\_bal:\*  
3: gb\_bal:\*  
4: gb\_bal:\*  
5: gb\_bal:\*  
6: gb\_bal:\*  
7: gb\_bal:\*  
8: gb\_bal:\*  
9: gb\_bal:\*  
10: gb\_bal:\*  
11: gb\_bal:\*  
12: gb\_bal:\*  
13: gb\_bal:\*  
14: gb\_bal:\*  
15: gb\_bal:\*  
16: gb\_bal:\*  
17: gb\_bal:\*  
18: gb\_bal:\*  
19: gb\_bal:\*  
20: gb\_bal:\*  
21: gb\_bal:\*  
22: gb\_bal:\*  
23: gb\_bal:\*  
24: gb\_bal:\*  
25: gb\_bal:\*  
26: gb\_bal:\*  
27: gb\_bal:\*  
28: gb\_bal:\*  
29: gb\_bal:\*  
30: gb\_bal:\*  
31: gb\_bal:\*  
32: gb\_bal:\*  
33: gb\_bal:\*  
34: gb\_bal:\*  
35: gb\_bal:\*  
36: gb\_bal:\*  
37: gb\_bal:\*  
38: gb\_bal:\*  
39: gb\_bal:\*  
40: gb\_bal:\*  
41: gb\_bal:\*  
42: gb\_bal:\*  
43: gb\_bal:\*  
44: gb\_bal:\*

45: gb\_bal:\*  
46: gb\_bal:\*  
47: gb\_bal:\*  
48: gb\_bal:\*  
49: gb\_bal:\*  
50: gb\_bal:\*  
51: gb\_bal:\*  
52: gb\_bal:\*  
53: gb\_bal:\*  
54: gb\_bal:\*  
55: gb\_bal:\*  
56: gb\_bal:\*  
57: gb\_bal:\*  
58: gb\_bal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1234	99.9	1235	5	A87609 Sequence 6
2	1234	99.9	1235	5	A89749 Sequence 6
3	1232.4	99.8	1235	5	A87611 Sequence 8
4	1232.4	99.8	1235	5	A89751 Sequence 8
5	1225.4	99.2	1229	5	A87610 Sequence 7
6	1225.4	99.2	1229	5	A89750 Sequence 7
7	1223	99.0	1236	5	A87613 Sequence 10
8	1223	99.0	1236	5	A89753 Sequence 10
9	1221.4	98.9	1237	5	A87607 Sequence 4
10	1221.4	98.9	1237	5	A89747 Sequence 4
11	1221.4	98.9	56414	1	MTU002
12	1219	98.7	1235	5	A87606 Sequence 3
13	1219	98.7	1235	5	A89746 Sequence 3
14	1213.4	98.3	1228	5	A87608 Sequence 5
15	1213.4	98.3	1228	5	A89748 Sequence 5
16	1207	97.7	1245	5	A87605 Sequence 2
17	1207	97.7	1245	5	A89745 Sequence 2
18	1207	97.7	1260	5	A87604 Sequence 1
19	1207	97.7	1260	5	A89744 Sequence 1
20	1207	97.7	2584	1	MTALADH
21	1195	96.8	1209	5	A87612 Sequence 9
22	1188	96.2	1208	5	A89752 Sequence 9
23	1182	95.7	1194	2	MTU92472
24	686	55.5	720	5	A89759
25	526.4	42.6	40745	1	SC151
26	455.6	36.9	682	5	A89758
27	455.6	36.9	682	5	A89758
28	365	29.6	8046	2	AF049107
29	340.8	27.6	11514	2	AE002028
30	319.4	25.9	2780	1	RLE238118
31	318.8	25.8	1585	2	AF070716
32	302.4	24.5	1736	1	BACALDHC
33	281	22.8	1477	2	AF070715
34	278	22.5	2113	5	E01847
35	266.8	21.6	1125	5	E01847
36	249.2	20.2	137740	1	D90900
37	232.8	18.9	1375	1	BACALDH
38	231	18.7	1776	1	BACALDH
39	225.2	18.2	1709	1	BACALDH
40	225.2	18.2	1709	1	BACALDH
41	225.2	18.2	1709	1	BACALDH
42	207.8	16.8	217420	2	AF070714
43	100	8.1	10552	2	AE001565
44	92	7.4	10512	2	AE000639
45	74	6.0	3160	1	RR005294

## ALIGNMENTS

```

RESULT 1
LOCUS A87609 1235 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 6 from Patent WO9836089.
ACCESSION A87609
VERSION A87609.1 GI:6736249
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1235)
AUTHORS Flohe, L. and Singh, M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
Source 1..1235
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 235 a 395 c 384 g 220 t 1 others
ORIGIN
Query Match 99.9%; Score 1234; DB 5; Length 1235;
Best Local Similarity 100.0%; Pred. No. 6,8e-167;
Matches 1235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTTGAGATTATGCAACTTTCTTCACTGAGAGCTACAGTATCGAGGGGTAATC 60
DB 1 ATCTTGAGATTATGCAACTTTCTTCACTGAGAGCTACAGTATCGAGGGGTAATC 60
QY 61 ATGCGCGTGTGATTTCGACGAGACCAAAACAAGATTCCGGGTGCCATCACCCG 120
DB 61 ATGCGCGTGTGATTTCGACGAGACCAAAACAAGATTCCGGGTGCCATCACCCG 120
QY 121 GCCGGCGTGGGAGACTAACCCGTGTGCGCATGAGTGTCATCAGCAGGTGCCGA 180
DB 121 GCCGGCGTGGGAGACTAACCCGTGTGCGCATGAGTGTCATCAGCAGGTGCCGA 180
QY 122 GCCGGCGTGGGAGACTAACCCGTGTGCGCATGAGTGTCATCAGCAGGTGCCGA 180
DB 122 GCCGGCGTGGGAGACTAACCCGTGTGCGCATGAGTGTCATCAGCAGGTGCCGA 180
QY 181 GAGGCGTGGCTATACCGACGCGGATTTCAAGGCGGCGCAATGCTGGGACG 240
DB 181 GAGGCGTGGCTATACCGACGCGGATTTCAAGGCGGCGCAATGCTGGGACG 240
QY 181 GAGGCGTGGCTATACCGACGCGGATTTCAAGGCGGCGCAATGCTGGGACG 240
DB 181 GAGGCGTGGCTATACCGACGCGGATTTCAAGGCGGCGCAATGCTGGGACG 240
QY 241 GCCGACAGGTGTGGCGGACGCTATTATGCTCAAGCTCAAGACCATACGGCG 300
DB 241 GCCGACAGGTGTGGCGGACGCTATTATGCTCAAGCTCAAGACCATACGGCG 300
QY 241 GCCGACAGGTGTGGCGGACGCTATTATGCTCAAGCTCAAGACCATACGGCG 300
DB 241 GCCGACAGGTGTGGCGGACGCTATTATGCTCAAGCTCAAGACCATACGGCG 300
QY 301 GAATAGGCGGCTGTGACAGACGCGGATTTGTCAGCTTCTTGATTTGGCGGTCAC 360
DB 301 GAATAGGCGGCTGTGACAGACGCGGATTTGTCAGCTTCTTGATTTGGCGGTCAC 360
QY 361 GTGCTTGACGAGTGTGTTGATTCGCGACACGTCATTTCCCTACAGACCGTCC 420
DB 361 GTGCTTGACGAGTGTGTTGATTCGCGACACGTCATTTCCCTACAGACCGTCC 420
QY 361 GTGCTTGACGAGTGTGTTGATTCGCGACACGTCATTTCCCTACAGACCGTCC 420
DB 361 GTGCTTGACGAGTGTGTTGATTCGCGACACGTCATTTCCCTACAGACCGTCC 420
QY 421 AAGACGCGGACGCGACACTACCCCTGCTTCCCGCATGAGCGAAGTCCGGTCCACTCG 480
DB 421 AAGACGCGGACGCGACACTACCCCTGCTTCCCGCATGAGCGAAGTCCGGTCCACTCG 480
QY 481 CCGCCAGGTTGGCGCTTACCACTGATGCAACCAAGGGGCGCGGTGTGCTGATGG 540
DB 481 CCGCCAGGTTGGCGCTTACCACTGATGCAACCAAGGGGCGCGGTGTGCTGATGG 540
QY 481 CCGCCAGGTTGGCGCTTACCACTGATGCAACCAAGGGGCGCGGTGTGCTGATGG 540
DB 481 CCGCCAGGTTGGCGCTTACCACTGATGCAACCAAGGGGCGCGGTGTGCTGATGG 540
QY 541 GCGGGGTGCGGCGTTCGAACGCGCGGACGCTGTGATTCGCGCGCGGACCGCGGT 600
DB 541 GCGGGGTGCGGCGTTCGAACGCGCGGACGCTGTGATTCGCGCGCGGACCGCGGT 600
QY 541 GCGGGGTGCGGCGTTCGAACGCGCGGACGCTGTGATTCGCGCGCGGACCGCGGT 600
DB 541 GCGGGGTGCGGCGTTCGAACGCGCGGACGCTGTGATTCGCGCGCGGACCGCGGT 600
QY 601 ACAAGCAGCGCGCATGCGACGCGCATGGGCGGACGCTTACGCTTACGACATCAACA 660
DB 601 ACAAGCAGCGCGCATGCGACGCGCATGGGCGGACGCTTACGCTTACGACATCAACA 660
QY 601 ACAAGCAGCGCGCATGCGACGCGCATGGGCGGACGCTTACGCTTACGACATCAACA 660
DB 601 ACAAGCAGCGCGCATGCGACGCGCATGGGCGGACGCTTACGCTTACGACATCAACA 660
QY 661 TCGACAAACTTGGCAACTCGACGCGGAGTTGTGCGCGGATCCACACTCGCTACTCAT 720

```

```

DB 661 TCGACAAACTTGGCAACTCGACGCGGAGTTGTGCGCGGATCCACACTCGCTACTCAT 720
QY 721 CGGCGTACAGAGTTCGAGGTTGCGTCAAAACGTCGACCTGGTGTGGGCGCGCTCGG 780
DB 721 CGGCGTACAGAGTTCGAGGTTGCGTCAAAACGTCGACCTGGTGTGGGCGCGCTCGG 780
QY 781 TGGCAGGCGCCAAAGCACCACCAATTAGTCTCGAATTCATTTGCGGATATGAACACG 840
DB 781 TGGCAGGCGCCAAAGCACCACCAATTAGTCTCGAATTCATTTGCGGATATGAACACG 840
QY 841 GTGCGGTACTGTGATATAGCATATGACATGACAGGCGGCGCTTTTCAAGCTCACAGCGA 900
DB 841 GTGCGGTACTGTGATATAGCATATGACATGACAGGCGGCGCTTTTCAAGCTCACAGCGA 900
QY 901 CCACCTACGACACCCGACGAGTTCGCGCTGACAGACAGCGTGTCTTACTGCGTGGCAACA 960
DB 901 CCACCTACGACACCCGACGAGTTCGCGCTGACAGACAGCGTGTCTTACTGCGTGGCAACA 960
QY 961 TGGCGCGCTGTGCGGCGGAGAGAGCTGACCTACCGCTGACCAACGCGAGATGCCGTATG 1020
DB 961 TGGCGCGCTGTGCGGCGGAGAGAGCTGACCTACCGCTGACCAACGCGAGATGCCGTATG 1020
QY 1021 TGGCTGAGCTTGGCGACATGAGCTGGGCGGCGGCGTCCGATCCGCGACTATAGCA 1080
DB 1021 TGGCTGAGCTTGGCGACATGAGCTGGGCGGCGGCGTCCGATCCGCGACTATAGCA 1080
QY 1081 AAGGCTTTTCGACGACGAAAGGGGCGCTTACTGTCCACAGGCTGGCGACGCTGGGGG 1140
DB 1081 AAGGCTTTTCGACGACGAAAGGGGCGCTTACTGTCCACAGGCTGGCGACGCTGGGGG 1140
QY 1141 TGGCGTTACAGGCGCGCGGACGCTGCTGCGCATCTCGGCGCGCTGCTTACGCGGAC 1200
DB 1141 TGGCGTTACAGGCGCGCGGACGCTGCTGCGCATCTCGGCGCGCTGCTTACGCGGAC 1200
QY 1201 ACAAGTGGGAGTAAAGGAAAGCATGATGTGGGCC 1235
DB 1201 ACAAGTGGGAGTAAAGGAAAGCATGATGTGGGCC 1235

RESULT 2
LOCUS A89749 1235 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 6 from Patent WO9832862.
ACCESSION A89749
VERSION A89749.1 GI:6738283
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1235)
AUTHORS Flohe, L. and Singh, M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
Source 1..1235
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 235 a 395 c 384 g 220 t 1 others
ORIGIN
Query Match 99.9%; Score 1234; DB 5; Length 1235;
Best Local Similarity 100.0%; Pred. No. 6,8e-167;
Matches 1235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTTGAGATTATGCAACTTTCTTCACTGAGAGCTACAGTATCGAGGGGTAATC 60
DB 1 ATCTTGAGATTATGCAACTTTCTTCACTGAGAGCTACAGTATCGAGGGGTAATC 60
QY 61 ATGCGCGTGTGATTTCGACGAGACCAAAACAAGATTCCGGGTGCCATCACCCG 120
DB 61 ATGCGCGTGTGATTTCGACGAGACCAAAACAAGATTCCGGGTGCCATCACCCG 120

```



```

Db 61 ATGCGGTGCGTATTCGACCGAGACAAACAAAGATTCGGGTGCGCATCACCCG 120
Oy 121 GCCGGGTGCGGAACTAACCCCTGCTGCGCATGAGTCTCATCGAGCAGTGC 180
Oy 121 GCCGGGTGCGGAACTAACCCCTGCTGCGCATGAGTCTCATCGAGCAGTGC 180
Db 121 GCCGGGTGCGGAACTAACCCCTGCTGCGCATGAGTCTCATCGAGCAGTGC 180
Oy 181 GAGGGGTGCGGATATCCGACGAGGATTCAGAGGGGCGGCGCAACTGCTGCGCAC 240
Oy 181 GAGGGGTGCGGATATCCGACGAGGATTCAGAGGGGCGGCGCAACTGCTGCGCAC 240
Db 181 GAGGGGTGCGGATATCCGACGAGGATTCAGAGGGGCGGCGCAACTGCTGCGCAC 240
Oy 241 GCGGACGAGTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Oy 241 GCGGACGAGTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Db 241 GCGGACGAGTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Oy 301 GAATACGCGCGCTGCGACACGAGGCGCATCTGTCTACGCTTCTTGCATTTGGCGGCTAC 360
Oy 301 GAATACGCGCGCTGCGACACGAGGCGCATCTGTCTACGCTTCTTGCATTTGGCGGCTAC 360
Db 301 GAATACGCGCGCTGCGACACGAGGCGCATCTGTCTACGCTTCTTGCATTTGGCGGCTAC 360
Oy 361 GTGCTTGACACGATGCGTGTGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Oy 361 GTGCTTGACACGATGCGTGTGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Db 361 GTGCTTGACACGATGCGTGTGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Oy 421 AGACCGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Oy 421 AGACCGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Db 421 AGACCGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Oy 481 CCGGCGGACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
Oy 481 CCGGCGGACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
Db 481 CCGGCGGACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
Oy 541 GCGGGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
Oy 541 GCGGGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
Db 541 GCGGGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
Oy 601 ACAACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
Oy 601 ACAACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
Db 601 ACAACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
Oy 661 TCGACAACTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
Oy 661 TCGACAACTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
Db 661 TCGACAACTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
Oy 721 CCGGCTACGAGCTGCGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
Oy 721 CCGGCTACGAGCTGCGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
Db 721 CCGGCTACGAGCTGCGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
Oy 781 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
Oy 781 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
Db 781 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
Oy 841 GTGCGTACTGCTGATATGAGCATGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
Oy 841 GTGCGTACTGCTGATATGAGCATGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
Db 841 GTGCGTACTGCTGATATGAGCATGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
Oy 901 CCACCTACGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
Oy 901 CCACCTACGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
Db 901 CCACCTACGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
Oy 961 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
Oy 961 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
Db 961 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
Oy 1021 TCGTGAAGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
Oy 1021 TCGTGAAGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
Db 1021 TCGTGAAGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
Oy 1081 AAGGCTTTGACGACGAGAGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
Oy 1081 AAGGCTTTGACGACGAGAGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
Db 1081 AAGGCTTTGACGACGAGAGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
Oy 1141 TCGCGTTACCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
Oy 1141 TCGCGTTACCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
Db 1141 TCGCGTTACCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200

```

```

Oy 1201 ACACGTGCGGAGTAAGGAGAGGATGATGCGCC 1235
Db 1201 ACACGTGCGGAGTAAGGAGAGGATGATGCGCC 1235

RESULT 3
A87611 1235 bp DNA PAT 22-JAN-2000
LOCUS A87611
DEFINITION Sequence 8 from Patent NO9836089.
ACCESSION A87611
VERSION A87611.1 GI:6736251
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1235)
AUTHORS Flohe, L. and Singh, M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
source
1..1235
location/qualifiers
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 394 c 385 g 220 t
ORIGIN

Query Match 99.8%; Score 1232.4; Df 5; Length 1235;
Best Local Similarity 99.8%; Pred. No. 1.le-16;
Matches 1233; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ATCTTGACGATTAATGCAACTTCTTACACTGAGGCGTACGATGAGAGGGGTAATC 60
Db 1 ATCTTGACGATTAATGCAACTTCTTACACTGAGGCGTACGATGAGAGGGGTAATC 60
Oy 61 ATGCGCGTGGTATTCCGACGAGCAAAACAAACGAATTCGGGGTCCATACCCCG 120
Oy 61 ATGCGCGTGGTATTCCGACGAGCAAAACAAACGAATTCGGGGTCCATACCCCG 120
Db 61 ATGCGCGTGGTATTCCGACGAGCAAAACAAACGAATTCGGGGTCCATACCCCG 120
Oy 121 GCGGCGTGGGCGAATACCCGCTGCGCATGAGGCGCATGAGGCGGCGGCGGCGG 180
Oy 121 GCGGCGTGGGCGAATACCCGCTGCGCATGAGGCGCATGAGGCGGCGGCGGCGGCGG 180
Db 121 GCGGCGTGGGCGAATACCCGCTGCGCATGAGGCGCATGAGGCGGCGGCGGCGGCGG 180
Oy 181 GAGGGCTGGGCTATCAGCAGCGGATTTCAAGGCGGCGAGGCGGCGCACTGGTGGCAC 240
Oy 181 GAGGGCTGGGCTATCAGCAGCGGATTTCAAGGCGGCGAGGCGGCGCACTGGTGGCAC 240
Db 181 GAGGGCTGGGCTATCAGCAGCGGATTTCAAGGCGGCGAGGCGGCGCACTGGTGGCAC 240
Oy 241 GCGGACGAGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Oy 241 GCGGACGAGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Db 241 GCGGACGAGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Oy 301 GAATACGCGCGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
Oy 301 GAATACGCGCGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
Db 301 GAATACGCGCGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
Oy 361 GTGCTTGACACGATGCGTGTGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Oy 361 GTGCTTGACACGATGCGTGTGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Db 361 GTGCTTGACACGATGCGTGTGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Oy 421 AGACCGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Oy 421 AGACCGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Db 421 AGACCGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Oy 481 CCGGCGGAGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
Oy 481 CCGGCGGAGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
Db 481 CCGGCGGAGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
Oy 541 GCGGGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
Oy 541 GCGGGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
Db 541 GCGGGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600

```

QY	601	ACAGCGACC	CCGCACTTCGGCAAGCGATGGGGCGCAGCCGTTCACGGTTTACAGCATCAACA	660
Db	601	ACAGCGACC	CCGCACTTCGGCAAGCGATGGGGCGCAGCCGTTCACGGTTTACAGCATCAACA	660
QY	661	TGCAGAAATTGGGCAACTCGAAGCGCGAATTTGTGCGGCCGGATTCACATCTCCTACTCAT		720
Db	661	TGCAGAAATTGGGCAACTCGAAGCGCGAATTTGTGCGGCCGGATTCACATCTCCTACTCAT		720
QY	721	CGGCGTACGAGCTCGAAGGTTGCCGTCATAAACGTCGCAGACTGGTGATGGGGCCGTCCTGG		780
Db	721	CGGCGTACGAGCTCGAAGGTTGCCGTCATAAACGTCGCAGACTGGTGATGGGGCCGTCCTGG		780
QY	781	TGCCAGGCGCCAAAGGCAACCCAATTAGTGTCTGAATTCACTGTGCGGCATATGAACAACAG		840
Db	781	TGCCAGGCGCCAAAGGCAACCCAATTAGTGTCTGAATTCACTGTGCGGCATATGAACAACAG		840
QY	841	GTCGGGTACTGGTGGATATAGCCATCGACCAAGGGGGGCTGTTTCAGAAGCTCACAGCGA		900
Db	841	GTCGGGTACTGGTGGATATAGCCATCGACCAAGGGGGGCTGTTTCAGAAGCTCACAGCGA		900
QY	901	CCACCTACGACACCGCAGCTTCGCGCGTGCAGACAGACGGTGTGTTTACTCGTGGCGAACA		960
Db	901	CCACCTACGACACCGCAGCTTCGCGCGTGCAGACAGACGGTGTGTTTACTCGTGGCGAACA		960
QY	961	TGCCCGGCTCGGTGCCGAAGACGTGAGCTTAGCGGCTGAACCAACGCGAGATGCCGTATG		1020
Db	961	TGCCCGGCTCGGTGCCGAAGACGTGAGCTTAGCGGCTGAACCAACGCGAGATGCCGTATG		1020
QY	1021	TGCTGAGAGTTCCCGACATATGGCTGGCGGGCGCGCGTGCAGATTCGGGACTAGGCA		1080
Db	1021	TGCTGAGAGTTCCCGACATATGGCTGGCGGGCGCGCGTGCAGATTCGGGACTAGGCA		1080
QY	1081	AAGGCTTTTCAGACGACGAAGGGGGGCTTACTGTCCGAACGGGTGGGCCACCGACTGGGGG		1140
Db	1081	AAGGCTTTTCAGACGACGAAGGGGGGCTTACTGTCCGAACGGGTGGGCCACCGACTGGGGG		1140
QY	1141	TGCGGTTACCGACGCCCCGCGACAGCGTGCCTGAGTCTCGGCGCGCTGCTTACGCCGANC		1200
Db	1141	TGCGGTTACCGACGCCCCGCGACAGCGTGCCTGAGTCTCGGCGCGCTGCTTACGCCGAGC		1200
QY	1201	ACACGTCGGGAGTAAGGGAAGCATGATGTCGGCC		1235
Db	1201	ACACGTCGGGAGTAAGGGAAGCATGATGTCGGCC		1235
RESULT	4			
LOCUS	A89751	1235 bp	DNA	PAT
DEFINITION	Sequence 8 from Patent WO9832862.			22-JAN-2000
ACCESSION	A89751			
VERSION	A89751.1	GI:6738285		
KEYWORDS				
SOURCE		unidentified.		
ORGANISM		unclassified.		
REFERENCE		1 (bases 1 to 1235)		
AUTHORS		Flohe L. and Singh M.		
TITLE		L-LANTINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM		
JOURNAL		Patent: WO 9832862-A 30-JUL-1998;		
FEATURES		FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)		
source		Location/Qualifiers		
		1..1235		
		/organism="unidentified"		
		/db_xref="taxon:32644"		
BASE COUNT	236 a	394 c	385 g	220 t
ORIGIN				
Query Match		99.8%	Score 1232.4;	DB 5; Length 1235;
Best Local Similarity		99.8%;	Pred. No. 1,1e+166;	
Matches 1233; Conservative		0;	Mismatches 2;	Indels 0; Gaps 0;
1 ATCTTGCAGATTAAATCGAATTCTTCACTGAAGCGTACAGATATCGAGAGGGTAATC				60

Db	1	ATCTTGACGATTAAATCGAACTTCTTCACTCAACGTACACATATCGAGGGGTAAATC	60
Qy	61	ATGCGCGTCGGTATATTCGACACGAGCAACAAAACAAAGAAATTCGGGGTGGCATACACCG	120
Db	61	ATGGCGTCGGGTATTTCGACCGCAGCAACCAAAAACAAAGAAATTCGGGGTGGCATACACCG	120
Qy	121	GCGGCGTCGCGGAATPACCCTGCTGGCCATAGAGTGTCTATCCAGCCAGGTCCGGA	180
Db	121	GCGGCGTCGCGGAATPACCCTGCTGGCCATAGAGTGTCTATCCAGCCAGGTCCGGA	180
Qy	181	GAGGCGTCGGCTATCCAGCAGCGCGATTTTAAAGCGCAGCGCCGCACTGGTCGGCAAC	240
Db	181	GAGGCGTCGGCTATCCAGCAGCGCGATTTTAAAGCGCAGCGCCGCACTGGTCGGCAAC	240
Qy	241	GCCGACCAAGTGTGGGCCGACGCTGATTTATGCTCAAGGCTAAAGAACCGATAGCGCG	300
Db	241	GCCGACCAAGTGTGGGCCGACGCTGATTTATGCTCAAGGCTAAAGAACCGATAGCGCG	300
Qy	301	GAAATCGGCGCCCTGTGCGCAAGGGCGATCTGTTCACGTTCTGGCATTTTGGCGGGTAC	360
Db	301	GAAATCGGCGCCCTGTGCGCAAGGGCGATCTGTTCACGTTCTGGCATTTTGGCGGGTAC	360
Qy	361	GTCGTTGCACCGATGCGTGTGTGGATTCCGGCACACGCTCAATTGGCTACGAGACGTC	420
Db	361	GTCGTTGCACCGATGCGTGTGTGGATTCCGGCACACGCTCAATTGGCTACGAGACGTC	420
Qy	421	AGACGGCGGAGGGGCACTACCCCTGTTCCCGCATGAGCCAAAGTCGCCGGTCCACATCG	480
Db	421	AGACGGCGGAGGGGCACTACCCCTGTTCCCGCATGAGCCAAAGTCGCCGGTCCACATCG	480
Qy	481	CCGCCAGGTTGGGCTTACACCTGATTCGGAACCCAAAGGGGCGCGGGTGTGTGATGG	540
Db	481	CCGCCAGGTTGGGCTTACACCTGATTCGGAACCCAAAGGGGCGCGGGTGTGTGATGG	540
Qy	541	GCGGGGTCGCCGCTGGAACCGCGCGAGCTGTGTATCGGCGCCGACCCGCGCT	600
Db	541	GCGGGGTCGCCGCTGGAACCGCGCGAGCTGTGTATCGGCGCCGACCCGCGCT	600
Qy	601	ACAAGCGAGCCCGATCGCCAAAGGGATGGGGCGCAGCGTTACGGTTTAAAGATCAACA	660
Db	601	ACAAGCGAGCCCGATCGCCAAAGGGATGGGGCGCAGCGTTTACGGTTTAAAGATCAACA	660
Qy	661	TCGACAAATTCGGCAATCGACCGCGAGTTCTGGCGCGGATCCACTGCTCTACTAT	720
Db	661	TCGACAAATTCGGCAATCGACCGCGAGTTCTGGCGCGGATCCACTGCTCTACTAT	720
Qy	721	GCGGCTACGACCTGAGAGGTGCTCAACAGTCCCACTGGTGAATGGGCGCGTCTGG	780
Db	721	GCGGCTACGACCTGAGAGGTGCTCAACAGTCCCACTGGTGAATGGGCGCGTCTGG	780
Qy	781	TGCCAGGCGCAAGGCAACCAATATCTCGAATTCACCTGTCGGCGATATGAACACAG	840
Db	781	TGCCAGGCGCAAGGCAACCAATATCTCGAATTCACCTGTCGGCGATATGAACACAG	840
Qy	841	GTCGGGTACTGTGGATATAGCATGACACGAGGGCGCTGTTCGAAGGCTACGACCGA	900
Db	841	GTCGGGTACTGTGGATATAGCATGACACGAGGGCGCTGTTCGAAGGCTACGACCGA	900
Qy	901	CCACCTACGACACCCGACGTCGCGGTGAGACACACGCGTTTAACTCGCGTGGGGAACA	960
Db	901	CCACCTACGACACCCGACGTCGCGGTGAGACACACGCGTTTAACTCGCGTGGGGAACA	960
Qy	961	TGCCGCTCGGTGCGAAGACGTGACCTACGCGCTGACCAACGGGAGAGTCCGTATG	1020
Db	961	TGCCGCTCGGTGCGAAGACGTGACCTACGCGCTGACCAACGGGAGAGTCCGTATG	1020
Qy	1021	TGCTGAGACTTGGCGACCATATGCGTGGGGGGCGCGTCCGCTGCAATCCGGCACTAGCA	1080
Db	1021	TGCTGAGACTTGGCGACCATATGCGTGGGGGGCGCGTCCGCTGCAATCCGGCACTAGCA	1080
Qy	1081	AAGCTCTTCAGCACGAGAGGGGCGTACTGTGCCAAAGGGGGCACGACCTGGGGG	1140

Db 1081 AAGGCTTTGACGACGACGAGGAGGAGGCTTACTGTCCGAAAGGGGTGGCCACCGACCTGGGGG 1140  
 QY 1141 TGCCGTTACGAGAGCCGCCGACGCTGTGGCTGTACTCTGGCGGCTGTTACGCCGANC 1200  
 Db 1141 TGCCGTTACGAGAGCCGCCGACGCTGTGGCTGTACTCTGGCGGCTGTTACGCCGAGC 1200  
 QY 1201 ACAGCTGGAGATAGGAGGAGGATGTCGGCC 1235  
 Db 1201 ACAGCTGGAGATAGGAGGAGGATGTCGGCC 1235

RESULT 5  
 A87610 1229 bp DNA PAT 22-JAN-2000  
 LOCUS A87610  
 DEFINITION Sequence 7 from Patent WO9836089.  
 ACCESSION A87610  
 VERSION A87610.1 GI:6736250  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1229)  
 AUTHORS Flohe, L. and Singh, M.  
 TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source 1..1229  
 /organism="unidentified"  
 /db\_xref="taxon:32644"

BASE COUNT 236 a 391 c 382 g 219 t 1 others  
 ORIGIN

Query Match 99.2%; Score 1225.4; DB 5; Length 1229;  
 Best Local Similarity 99.8%; Pred. No. 1.1e-165;  
 Matches 1226; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ANCTTCGAGATTATGCACTTTCTTACACATGTAAGCTGATGAGAGGGGTATC 60  
 Db 1 ANCTTCGAGATTATGCACTTTCTTACACATGTAAGCTGATGAGAGGGGTATC 60  
 QY 61 ATGCGGCTGGATTTCGACGACGAGACCAAAACAAAGATTCGGGGTGGCCATCACC 120  
 Db 61 ATGCGGCTGGATTTCGACGACGAGACCAAAACAAAGATTCGGGGTGGCCATCACC 120  
 QY 121 GCGGCGTGGCGAATTAACCCGCTGCGCATGAGTGTCTATCCAGGACAGTGGCGGA 180  
 Db 121 GCGGCGTGGCGAATTAACCCGCTGCGCATGAGTGTCTATCCAGGACAGTGGCGGA 180  
 QY 181 GAGGCTCGGCTATACCGACGCGGATTTCAAGGGGCGAGCGGCACTGGTGGCACC 240  
 Db 181 GAGGCTCGGCTATACCGACGCGGATTTCAAGGGGCGAGCGGCACTGGTGGCACC 240  
 QY 241 GCGGACGAGTGGGCGGCGGCGCTGATTTATGCTCAAGGTCAAAGACGATAGGGGG 300  
 Db 241 GCGGACGAGTGGGCGGCGGCGCTGATTTATGCTCAAGGTCAAAGACGATAGGGGG 300  
 QY 301 GAATACGGCGCTGCGACACGAGGCGATTTCTTCAAGTTCCTTGGCGGCTCAC 360  
 Db 301 GAATACGGCGCTGCGACACGAGGCGATTTCTTCAAGTTCCTTGGCGGCTCAC 360  
 QY 361 GTGCTTGACACGATGCTTTGTGATTCGCGACACGATTAATGGCTACGAGACGCTGC 420  
 Db 361 GTGCTTGACACGATGCTTTGTGATTCGCGACACGATTAATGGCTACGAGACGCTGC 420  
 QY 421 AGACGCGGAGGCGGCGCTACCCCTGCTGCGGATGAGCGAAGTGGCGGCTGACTCG 480  
 Db 421 AGACGCGGAGGCGGCGCTACCCCTGCTGCGGATGAGCGAAGTGGCGGCTGACTCG 480  
 QY 481 CCGCCGAGGTTGGGCTTACACCTGATGGAACCCAGGGGGCGGCTGTCTGATGG 540  
 Db 481 CCGCCGAGGTTGGGCTTACACCTGATGGAACCCAGGGGGCGGCTGTCTGATGG 540

QY 541 GCGGGGTGCCCCGGCTGCAACCGGCGGACGTCGTGTGATCGGGCGGCGGCGGCT 600  
 Db 541 GCGGGGTGCCCCGGCTGCAACCGGCGGACGTCGTGTGATCGGGCGGCGGCGGCT 600  
 QY 601 ACAACGAGCCCGGATCGCCCAAGCGCATGGGCGGCGACCGCTTACGATTAACA 660  
 Db 601 ACAACGAGCCCGGATCGCCCAAGCGCATGGGCGGCGACCGCTTACGATTAACA 660  
 QY 661 TCGCAAACTTGGGCACTGAGCGCGGATTTGCGGCGCGGATTCACACCTGCTACTCAT 720  
 Db 661 TCGCAAACTTGGGCACTGAGCGCGGATTTGCGGCGCGGATTCACACCTGCTACTCAT 720  
 QY 721 CGGCTTACGAGCTGAGAGGTTGCGTCAAGCTGCGGACCTGTGTATGGGCGGCTCTCG 780  
 Db 721 CGGCTTACGAGCTGAGAGGTTGCGTCAAGCTGCGGACCTGTGTATGGGCGGCTCTCG 780  
 QY 781 TGCCAGGCGGCGAAGGACCCAAATTAGTCTGGAATTCTTGTGGCGATTTGAAACAG 840  
 Db 781 TGCCAGGCGGCGAAGGACCCAAATTAGTCTGGAATTCTTGTGGCGATTTGAAACAG 840  
 QY 841 GTGGGCTACTGTGATATATACCATGACGACGAGCGGCGTGTTCGAAGGCTCAGACCGA 900  
 Db 841 GTGGGCTACTGTGATATATACCATGACGACGAGCGGCGTGTTCGAAGGCTCAGACCGA 900  
 QY 901 CCACCTACGACACCCGCGAGCTTGGCGGTGACGACGACGCTGTTTACTGCGTGGGACA 960  
 Db 901 CCACCTACGACACCCGCGAGCTTGGCGGTGACGACGACGCTGTTTACTGCGTGGGACA 960  
 QY 961 TGCCGCGCTCGGTTGCGGAAGCTGTGACCTACGCGCTGACCAAGCGGATGGCGGATG 1020  
 Db 961 TGCCGCGCTCGGTTGCGGAAGCTGTGACCTACGCGCTGACCAAGCGGATGGCGGATG 1020  
 QY 1021 TGCTGAGCTTGGCGACCATGCTGGGCGGCGGCTGGCGGCTGCAATTCGCGACTAGCA 1080  
 Db 1021 TGCTGAGCTTGGCGACCATGCTGGGCGGCGGCTGGCGGCTGCAATTCGCGACTAGCA 1080  
 QY 1081 AAGGCTTTTCACGACGAGGCGGCTTACTGTCCGAAAGGCTGGCCACGACCTGGGG 1140  
 Db 1081 AAGGCTTTTCACGACGAGGCGGCTTACTGTCCGAAAGGCTGGCCACGACCTGGGG 1140  
 QY 1141 TGCCGTTACGAGAGCCGCCGACGCGTGTGCTGACCTGCGGCGCTGTTACGCCGAGC 1200  
 Db 1141 TGCCGTTACGAGAGCCGCCGACGCGTGTGCTGACCTGCGGCGCTGTTACGCCGAGC 1200  
 QY 1201 ACAGCTCGGAGATAGGAGGAGGATGATG 1229  
 Db 1201 ACAGCTCGGAGATAGGAGGAGGATGATG 1229

RESULT 6  
 A89750 1229 bp DNA PAT 22-JAN-2000  
 LOCUS A89750  
 DEFINITION Sequence 7 from Patent WO9832862.  
 ACCESSION A89750  
 VERSION A89750.1 GI:6738284  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1229)  
 AUTHORS Flohe, L. and Singh, M.  
 TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
 JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source 1..1229  
 /organism="unidentified"  
 /db\_xref="taxon:32644"

BASE COUNT 236 a 391 c 382 g 219 t 1 others  
 ORIGIN

Query Match 99.2%; Score 1225.4; DB 5; Length 1229;  
 Best Local Similarity 99.8%; Pred. No. 1.1e-165;  
 Matches 1229; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 ATCTTGAGATTAATGAACTTTCTTCACTGAAAGGTACAGATGAGAGGGGTATC 60
DB 1 ATCTTGAGATTAATGAACTTTCTTCACTGAAAGGTACAGATGAGAGGGGTATC 60
QY 61 ATGGGCGGTATTCGACGAGACCAAAACAAAGCAATCCGGTGGCCATCACCCG 120
DB 61 ATGGGCGGTATTCGACGAGACCAAAACAAAGCAATCCGGTGGCCATCACCCG 120
QY 121 GCGGCGGTGCGGAACTAACCCGTGTCGATGAGTGTCTATCCAGGAGGTGCGGA 180
DB 121 GCGGCGGTGCGGAACTAACCCGTGTCGATGAGTGTCTATCCAGGAGGTGCGGA 180
QY 181 GAGGCGGTGCTATCAGCAGGAGGCGGATTTCAAGGCGGAGCGCAACTGGTGGCACC 240
DB 181 GAGGCGGTGCTATCAGCAGGAGGCGGATTTCAAGGCGGAGCGCAACTGGTGGCACC 240
QY 241 GCGGACAGGTGTGGGCGGACGCTGATTTATGCTCAAGTCAAGAACGATAGCGGCG 300
DB 241 GCGGACAGGTGTGGGCGGACGCTGATTTATGCTCAAGTCAAGAACGATAGCGGCG 300
QY 301 GAATACGGCGCGCTGCGAGAGGCGGATCTGTTACAGCTT TGGCATTTGCGCGCTAC 360
DB 301 GAATACGGCGCGCTGCGAGAGGCGGATCTGTTACAGCTT TGGCATTTGCGCGCTAC 360
QY 361 GTGCTTGACAGCATGCTGTGTGATTCGCGACACGCTCAATTGCTTACAGACCGTCC 420
DB 361 GTGCTTGACAGCATGCTGTGTGATTCGCGACACGCTCAATTGCTTACAGACCGTCC 420
QY 421 AGACCGCGGAGCGGCGACATCCCTGCTTCCCGCATGAGGAAATCCCGCTGACTGCG 480
DB 421 AGACCGCGGAGCGGCGACATCCCTGCTTCCCGCATGAGGAAATCCCGCTGACTGCG 480
QY 481 CGGCCAGGTGGCGCTTACACCTGATGCGAACCAAGGGGCGCGGTGTGCTGATG 540
DB 481 CGGCCAGGTGGCGCTTACACCTGATGCGAACCAAGGGGCGCGGTGTGCTGATG 540
QY 541 GCGGGGTGCGGCGCTGCGAACCGGCGGAGCTGTGATTCGCGCGCGGACCGCGGCT 600
DB 541 GCGGGGTGCGGCGCTGCGAACCGGCGGAGCTGTGATTCGCGCGCGGACCGCGGCT 600
QY 601 ACAAGGCGGCGGCGATCGCGAACGAGGAGGCGGCGGCGCTTACGCTTACAGCATCA 660
DB 601 ACAAGGCGGCGGCGATCGCGAACGAGGAGGCGGCGGCGCTTACGCTTACAGCATCA 660
QY 661 TCGACAACTTGGGCACTGAGCGCGGAGTGTGCGCGGAG. ACACCTGCTACTCAT 720
DB 661 TCGACAACTTGGGCACTGAGCGCGGAGTGTGCGCGGAGTGTGCGCGGAGTGTGCG 720
QY 721 CGGCGTACGAGCTCGAGGGGCGGTCGAAAGCTGCGGAGTGTGCGGAGTGTGCGGAG 780
DB 721 CGGCGTACGAGCTCGAGGGGCGGTCGAAAGCTGCGGAGTGTGCGGAGTGTGCGGAG 780
QY 781 TCGCAGGCGGCAAGGCAACCAATAGTCTGCAATTCACCTTGTGCGGCAATGAAACAG 840
DB 781 TCGCAGGCGGCAAGGCAACCAATAGTCTGCAATTCACCTTGTGCGGCAATGAAACAG 840
QY 841 GTGCGGTATGCTGATGATTAAGCATGACAGGCGGCGCTTTCGAAAGGCTACAGCGGA 900
DB 841 GTGCGGTATGCTGATGATTAAGCATGACAGGCGGCGCTTTCGAAAGGCTACAGCGGA 900
QY 901 CCACCTACGACACCGGAGTGTGCGTACAGACGCTGTTTACTGCTGTCGCGGACGA 960
DB 901 CCACCTACGACACCGGAGTGTGCGTACAGACGCTGTTTACTGCTGTCGCGGACGA 960
QY 961 TGGCGGCTGCTGTCGAGAGAGCTGACCTACGCGCTGACCAAGCGGAGATGCCATG 1020
DB 961 TGGCGGCTGCTGTCGAGAGAGCTGACCTACGCGCTGACCAAGCGGAGATGCCATG 1020
QY 1021 TGCTCGAGCTTGGCGACCATGGGTGGGCGGCGGTGCGGATCGGCGACTAGCA 1080
  
```

```

DB 1021 TGCTCGAGCTTGGCGACCATGGGTGGGCGGCGGTGCGGATCGGCGACTAGCA 1080
QY 1081 AAGGCTTTTCAGCAGCAGAGAGGGGCTTACTGTCCGAACGGGTGGCCACGACCTGGGG 1140
DB 1081 AAGGCTTTTCAGCAGCAGAGAGGGGCTTACTGTCCGAACGGGTGGCCACGACCTGGGG 1140
QY 1141 TGGCGTTCACGAGCGCGGCGGAGGCTGTGCGCTGACTCTCGGCGCTGTTACGCCGANC 1200
DB 1141 TGGCGTTCACGAGCGCGGCGGAGGCTGTGCGCTGACTCTCGGCGCTGTTACGCCGAGC 1200
QY 1201 ACAGCTCGGAGTAAAGGAGGAGGATG 1229
DB 1201 ACAGCTCGGAGTAAAGGAGGAGGATG 1229
  
```

```

RESULT 7
A87613 1236 bp DNA PAT 22-JAN-2000
LOCUS Sequence 10 from Patent WO9836089.
DEFINITION A87613
ACCESSION A87613
VERSION A87613.1 GI:6736253
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1236)
AUTHORS Flohe L. and Singh M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-Aug-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
source
1..1236
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 395 c 385 g 220 t
ORIGIN
  
```

Query Match 99.0%; Score 1223; DB 5; Length 1236;  
 Best Local Similarity 99.8%; Pred. No. 2.5e-165;  
 Matches 1234; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```

QY 1 ATCTTGAGATTAATGAACTTTCTTCACTGAAAGGTACAGATGAGAGGGGTATC 60
DB 1 ATCTTGAGATTAATGAACTTTCTTCACTGAAAGGTACAGATGAGAGGGGTATC 60
QY 61 ATGGGCGGTATTCGACGAGACCAAAACAAAGCAATCCGGTGGCCATCACCCG 120
DB 61 ATGGGCGGTATTCGACGAGACCAAAACAAAGCAATCCGGTGGCCATCACCCG 120
QY 121 GCGGCGGTGCGGAACTAACCCGTGTCGATGAGTGTCTATCCAGGAGGTGCGGA 180
DB 121 GCGGCGGTGCGGAACTAACCCGTGTCGATGAGTGTCTATCCAGGAGGTGCGGA 180
QY 181 GAGGCGGTGCTATCAGCAGGAGGCGGATTTCAAGGCGGAGCGCAACTGGTGGCACC 240
DB 181 GAGGCGGTGCTATCAGCAGGAGGCGGATTTCAAGGCGGAGCGCAACTGGTGGCACC 240
QY 241 GCGGACAGGTGTGGGCGGACGCTGATTTATGCTCAAGTCAAGAACGATAGCGGCG 300
DB 241 GCGGACAGGTGTGGGCGGACGCTGATTTATGCTCAAGTCAAGAACGATAGCGGCG 300
QY 301 GAATACGGCGCGCTGCGAGAGGCGGATCTGTTACAGCTT TGGCATTTGCGCGCTAC 360
DB 301 GAATACGGCGCGCTGCGAGAGGCGGATCTGTTACAGCTT TGGCATTTGCGCGCTAC 360
QY 360 GGTGCTTACAGCATGCTGTGTGATTCGCGACACGCTCAATTTGCTTACGAGACGCTC 419
DB 361 GGTGCTTACAGCATGCTGTGTGATTCGCGACACGCTCAATTTGCTTACGAGACGCTC 420
QY 420 CAGACCGCGGAGCGGCGACTACCCCTGTTGCCCGATGAGCGGAAGTGGCGGTGCACTC 479
  
```

Db 421 CAGACCGCGACGGCGACTACCCCTGCTTGCCCGCATGAGCAAGTCCGGGTGCACTC 480  
 QY 480 GCGGCCAGGTTGGCGCTTACCACTGATGCGCAACCAAGGGGGCGGGTGTGCTGATG 539  
 Db 481 GCGGCCAGGTTGGCGCTTACCACTGATGCGCAACCAAGGGGGCGGGTGTGCTGATG 540  
 QY 540 GCGGGGGTGGCGGCTGCAAGCGGCGAGAGTGTGTGATGCGGGCGGCGACCGCGG 599  
 Db 541 GCGGGGGTGGCGGCTGCAAGCGGCGAGAGTGTGTGATGCGGGCGGCGACCGCGG 600  
 QY 600 TACACGCGACCGCGCATCGCCAAAGCGATGCGCGACCGCTTACGATGATCAAC 659  
 Db 601 TACACGCGACCGCGCATCGCCAAAGCGATGCGCGACCGCTTACGATGATCAAC 660  
 QY 660 ATGACAAATTGGGCACTGCAAGCGGAGTGTGTGCGGGCGGATCCACATCGCTACTCA 719  
 Db 661 ATGACAAATTGGGCACTGCAAGCGGAGTGTGTGCGGGCGGATCCACATCGCTACTCA 720  
 QY 720 TCGGCTACGAGCTCGAGGGTGGCGTCAACGTGCGGACCTGATGATGGGGCGGCTG 779  
 Db 721 TCGGCTACGAGCTCGAGGGTGGCGTCAACGTGCGGACCTGATGATGGGGCGGCTG 780  
 QY 780 GTGCGAGGCGCCAGGACCCAAATTAATGTCGAATTCACTGTGCGGCATATGAACA 839  
 Db 781 GTGCGAGGCGCCAGGACCCAAATTAATGTCGAATTCACTGTGCGGCATATGAACA 840  
 QY 840 GTGCGGCTACTGTGATATAGCATCGACAGGGGGCGCTTTGCAAGGCTCAACGAC 899  
 Db 841 GTGCGGCTACTGTGATATAGCATCGACAGGGGGCGCTTTGCAAGGCTCAACGAC 900  
 QY 900 ACCACTAGACACCGGAGCTGCGCGTGCAGACAGCGCTGTTTACTGCTGCGGAC 959  
 Db 901 ACCACTAGACACCGGAGCTGCGCGTGCAGACAGCGCTGTTTACTGCTGCGGAC 960  
 QY 960 ATGCGCGCTCGGTGCGGAAAGAGCTGACCTGACGCGCTGACCAAGCGGAGATCCGAT 1019  
 Db 961 ATGCGCGCTCGGTGCGGAAAGAGCTGACCTGACGCGCTGACCAAGCGGAGATCCGAT 1020  
 QY 1020 GTGCTGAGGTTGCCGACCATGCTGGCGGGCGGGCTGCGGCTGCAATCCGGCACTAGCC 1079  
 Db 1021 GTGCTGAGGTTGCCGACCATGCTGGCGGGCGGGCTGCGGCTGCAATCCGGCACTAGCC 1080  
 QY 1080 AAGGCTTTTCAGACGACGAAGGGGCGTTACTGTCCGAACGGGTGGCGACCGACTGGGG 1139  
 Db 1081 AAGGCTTTTCAGACGACGAAGGGGCGTTACTGTCCGAACGGGTGGCGACCGACTGGGG 1140  
 QY 1140 GTGCGGTTACCGAGCGCGCGACGCTGCTGCTGACTCTCGGCGCTCTTACGCGGAN 1199  
 Db 1141 GTGCGGTTACCGAGCGCGCGACGCTGCTGCTGACTCTCGGCGCTCTTACGCGGAN 1200  
 QY 1200 CACACGTGCGGAGTAAGGGAAGGATGATTCGGCC 1235  
 Db 1201 CACACGTGCGGAGTAAGGGAAGGATGATTCGGCC 1236

RESULT 8  
 A89753  
 LOCUS A89753 1236 bp DNA  
 DEFINITION Sequence 10 from Patient MO9832862.  
 ACCESSION A89753  
 VERSION A89753.1 GI:6738287  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 1 (bases 1 to 1236)  
 FLOHE, L. and Singh, M.  
 L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
 Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 Location/Qualifiers  
 1..1236  
 /organism="unidentified"

BASE COUNT 236 a 395 c 385 g 220 t  
 ORIGIN  
 Query Match 99.0%; Score 1223; DB 5; Length 1236;  
 Best local similarity 99.8%; Pred. No. 2,5e-165;  
 Matches 1234; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATCTTCAGATTAATGAACCTTTCTTACACTGAAACGCTACAGTATCGAAGGGGTAATC 60  
 Db 1 ATCTTCAGATTAATGAACCTTTCTTACACTGAAACGCTACAGTATCGAAGGGGTAATC 60  
 QY 61 ATGCGGCTGGGATATCCGACCGAGAACCAAAACAGAAATTCGGGGTGGCATACCCG 120  
 Db 61 ATGCGGCTGGGATATCCGACCGAGAACCAAAACAGAAATTCGGGGTGGCATACCCG 120  
 QY 121 GCGGCGTGGGAACTAACCCGCTCGTGGCCATGAGTGTCTATCCAGGAGAGTGGCGGA 180  
 Db 121 GCGGCGTGGGAACTAACCCGCTCGTGGCCATGAGTGTCTATCCAGGAGAGTGGCGGA 180  
 QY 181 GAGGCTTGGGCTATACCGACGCGGATTTCAAGCGCGAGCGCGCACTGGTGGCAC 240  
 Db 181 GAGGCTTGGGCTATACCGACGCGGATTTCAAGCGCGAGCGCGCACTGGTGGCAC 240  
 QY 241 GCGGACAGGTGGGGCCGACGCTGATTTATGCTCAAGCTCAAGAACCGATAGCGGG 300  
 Db 241 GCGGACAGGTGGGGCCGACGCTGATTTATGCTCAAGCTCAAGAACCGATAGCGGG 300  
 QY 301 GAATACGGCGCGCTGGGACGAGGAGG - GATCTGTTACGCTTGGCATTTGGCCGGGTCA 359  
 Db 301 GAATACGGCGCGCTGGGACGAGGAGGAGTCTGTTACGCTTGGCATTTGGCCGGGTCA 360  
 QY 360 CGTGTTCGACCGGATCGTTGTTGATTCGCGACACGCTCAATTGCTTACGAGAACCGTC 419  
 Db 361 CGTGTTCGACCGGATCGTTGTTGATTCGCGGACACGCTCAATTGCTTACGAGAACCGTC 420  
 QY 420 CAGACCGCGGAGGCGGCACTACCCCTGCTTGGCCGCGATAGAGAACTGCGCGGTGACTC 479  
 Db 421 CAGACCGCGGAGGCGGCACTACCCCTGCTTGGCCGCGATAGAGAACTGCGCGGTGACTC 480  
 QY 480 GCGGCGAGGTTGGCGCTTACCACTGATGCGCAACCAAGGGGGCGGGTGGCTGATG 539  
 Db 481 GCGGCGAGGTTGGCGCTTACCACTGATGCGCAACCAAGGGGGCGGGTGGCTGATG 540  
 QY 540 GCGGGGTTCCCGCGCTCGCAACCGGCGAGCTGCTGATGCGGCGCGCAACCGCGGCG 599  
 Db 541 GCGGGGTTCCCGCGCTCGCAACCGGCGAGCTGCTGATGCGGCGCGCAACCGCGGCG 600  
 QY 600 TACACGCGACCGCGCATTCGCGCAACGCGATGCGGCGGACGCTTACGCTTACGATCAAC 659  
 Db 601 TACACGCGACCGCGCATTCGCGCAACGCGATGCGGCGGACGCTTACGCTTACGATCAAC 660  
 QY 660 ATGACAAATTGGGCACTGCAAGCGGAGTGTGTGCGGGCGGATCCACATCGCTACTCA 719  
 Db 661 ATGACAAATTGGGCACTGCAAGCGGAGTGTGTGCGGGCGGATCCACATCGCTACTCA 720  
 QY 720 TCGGCTACGAGCTCGAGGGTGGCGTCAACGTGCGGACCTGATGATGGGGCGGCTG 779  
 Db 721 TCGGCTACGAGCTCGAGGGTGGCGTCAACGTGCGGACCTGATGATGGGGCGGCTG 780  
 QY 780 GTGCGAGGCGCCAGGACCCAAATTAATGTCGAATTCACTGTGCGGCATATGAACA 839  
 Db 781 GTGCGAGGCGCCAGGACCCAAATTAATGTCGAATTCACTGTGCGGCATATGAACA 840  
 QY 840 GTGCGGCTACTGTGATATAGCATCGACAGGGGGCGCTTTGCAAGGCTCAACGAC 899  
 Db 841 GTGCGGCTACTGTGATATAGCATCGACAGGGGGCGCTTTGCAAGGCTCAACGAC 900  
 QY 900 ACCACTAGACACCGGAGCTGCGCGTGCAGACAGCGCTGTTTACTGCTGCGGAGAC 959  
 Db 901 ACCACTAGACACCGGAGCTGCGCGTGCAGACAGCGCTGTTTACTGCTGCGGAGAC 960

Query Match	98.9%	Score 1221.4	DB: 5	Length 1237
Best Local Similarity	99.8%	Pred. No. 4,2e-165		
Matches 1233	Conservative	0	Mismatches 2	Indels 1
QY 1	ATCTTCGAGATTAAATCGAACTTCTTTCACACATCGAAGCGTACAGTATCGAGAGGGGTATTC	60		
Db 1	ATCTTCGAGATTAAATCGAACTTCTTTCATACGTGAAGCGTACAGTATCGAGAGGGGTATTC	60		
QY 61	ATGGCGGTGGTATTTCGACCGAGACCAAAAACAAAGAAATTCGGGTGGCATCACCCCG	120		
Db 61	ATGGCGGTGGTATTTCGACCGAGACCAAAAACAAAGAAATTCGGGTGGCATCACCCCG	120		
QY 121	GCCGGCGTGGCGGAATCAACCCGTGTGGCCATGAGGTCTCATCCAGGCAGAGTGGCCGA	180		
Db 121	GCCGGCGTGGCGGAATCAACCCGTGTGGCCATGAGGTCTCATCCAGGCAGAGTGGCCGA	180		
QY 181	GAGGGCTGGGCTATCACCGACGCGGATTTCAAGGGCGGAGGCGCGCACTGTGTGGCACC	240		
Db 181	GAGGGCTGGGCTATCACCGACGCGGATTTCAAGGGCGGAGGCGCGCACTGTGTGGCACC	240		
QY 241	GCCGACCAAGTGTGGGCGGACGCTGATTTATTGTCTCAAGGTCAAAAGAACCCGATAGCGGCG	300		
Db 241	GCCGACCAAGTGTGGGCGGACGCTGATTTATTGTCTCAAGGTCAAAAGAACCCGATAGCGGCG	300		
QY 301	GAAATACGGCGGCTCGACACAGGGGCGATCTGTTCACGTTCTTGATTTGGCCGGGTCA	359		
Db 301	GAAATACGGCGGCTCGACACAGGGGCGATCTGTTCACGTTCTTGATTTGGCCGGGTCA	360		
QY 350	CGTGGCTTCACCGAATGCGTGTGTGGATTCGGGACCAACCAATTCATTGCTTACGAGACCGTC	419		

RESULT	10			PAT	22-JAN-2000
A89747					
LOCUS	A89747	1237 bp	DNA		
DEFINITION	Sequence 4 from Patent WO9332862.				
ACCESSION	A89747				
VERSION	A89747.1	GI:6738281			
KEYWORDS	.				
SOURCE	unidentified.				
ORGANISM	unidentified.				
	unclassified.				
REFERENCE	1 (bases 1 to 1237)				
AUTHORS	Flohe,L. and Singh,M.				
TITLE	L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM				



JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES Location/Qualifiers  
 source 1.1237  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 236 a 394 c 386 g 221 t  
 ORIGIN

Query Match 98.9%; Score 1221.4; DB 5; Length 1237;  
 Best Local Similarity 99.8%; Pred. No. 4.2e-165;  
 Matches 1233; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

1 ATCTGAGATTATCGAATCTTCTTCACTGAGCGTACAGTATGAGAGGGTATC 60  
 1 ATCTGAGATTATCGAATCTTCTTCACTGAGCGTACAGTATGAGAGGGTATC 60  
 61 ATGCGCGTGGTATTCGACCGAGACCAAAACAAAGCAATTCGGGTGCGCATCCCG 120  
 61 ATGCGCGTGGTATTCGACCGAGACCAAAACAAAGCAATTCGGGTGCGCATCCCG 120  
 121 GCGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 180  
 121 GCGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 180  
 181 GAGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 240  
 181 GAGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 240  
 241 GCGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 300  
 241 GCGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 300  
 301 GAATACGCGCGCTGCGACGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 359  
 301 GAATACGCGCGCTGCGACGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 359  
 361 GCGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 420  
 361 GCGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 420  
 420 GCGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 480  
 420 GCGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 480  
 481 GCGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 540  
 481 GCGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 540  
 540 GCGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 599  
 540 GCGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 599  
 600 TACAAGGCAAGCGGATCGGCAAGCGGATCGGCAAGCGGATCGGCAAGCGGAT 659  
 600 TACAAGGCAAGCGGATCGGCAAGCGGATCGGCAAGCGGATCGGCAAGCGGAT 659  
 660 ATGACAAATTCGGAATCGGCAAGCGGATCGGCAAGCGGATCGGCAAGCGGAT 719  
 660 ATGACAAATTCGGAATCGGCAAGCGGATCGGCAAGCGGATCGGCAAGCGGAT 719  
 720 TCGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 779  
 720 TCGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 779  
 780 GTCGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 839  
 780 GTCGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 839  
 840 GTCGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 899  
 840 GTCGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 899

QY 900 ACCAGCTACGACGACCGGACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 959  
 DB 901 ACCAGCTACGACGACCGGACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960  
 QY 960 ATGCGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 1019  
 DB 961 ATGCGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 1020  
 QY 1020 GTCGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 1079  
 DB 1021 GTCGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 1080  
 QY 1080 AAGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 1139  
 DB 1081 AAGGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 1140  
 QY 1140 GTCGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 1199  
 DB 1141 GTCGCGTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 1200  
 QY 1200 CACAGCTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 1235  
 DB 1201 CACAGCTCCGGAACCTACCGTGGCATGAGGTGCTCATCCAGGAGGCGCGGA 1236

RESULT 11  
 MTW002 56414 bp DNA BCT 17-JUN-1998  
 LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.  
 DEFINITION AL008967 AL123456  
 ACCESSION AL008967.1 GI:3261491  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mycobacterium tuberculosis.  
 Mycobacterium tuberculosis.  
 Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Corynebacteriales; Mycobacteriaceae;  
 Mycobacterium.

REFERENCE  
 AUTHORS  
 Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,  
 Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,  
 Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,  
 Connor, R., Davies, R., Devlin, K., Felwell, T., Gentles, S.,  
 Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,  
 Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,  
 Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,  
 Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and  
 Barrell, B.G.

TITLE  
 JOURNAL  
 MEDLINE  
 REMARK  
 Erratum: [[published erratum appears in Nature 1998 Nov  
 12;396(6707):190]]  
 Nature 393 (6685), 537-544 (1998)  
 98295987

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium  
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome  
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique  
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,  
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
 On Jun 27, 1998 this sequence version replaced g12621256.  
 Notes:  
 Details of M. tuberculosis sequencing at the Sanger Centre are  
 available on the World Wide Web.  
 (URL, http://www.sanger.ac.uk/projects/M.tuberculosis/) CDS have  
 been renumbered from the original cosmid submissions so the old  
 gene designations are in brackets after the new gene numbers.  
 Gene prediction was based on a Hidden Markov Model of TB genes  
 implemented in Tsparse (Krogh) supplemented with visual inspection  
 of positional base preference in codons, especially where there is  
 an increase in the observed/expected third position G + C.

COMMENT

CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtc, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

## FEATURES

```

source      1..56414
            /organism="Mycobacterium tuberculosis"
            /strain="H37Rv"
            /db_xref="taxon:1773"
            <1..>56414
            /note="fragment designated v002. Does not represent a
            physical clone"
            <1..233
            /organism="Mycobacterium tuberculosis"
            /strain="H37Rv"
            /db_xref="taxon:1773"
            /clone="Y154"
            /complement(3..527)
            /gene="recX"
            /complement(3..527)
            /gene="recX"
            /note="RV2738c, (MTV002.01c), len: 174 aa. recX, similar
            to eg. RECX_PSEAE_P37860 regulatory protein recX from
            Pseudomonas aeruginosa (153 aa), fasta scores: opt: 161
            z-score: 257.2 E(): 3.6e-07, 30.7% identity in 137 aa
            overlap. Overlaps and extends CDS from overlapping cosmid
            MCY154.16c"
            /codon_start=1
            /transl_table=11
            /product="recX"
            /protein_id="CA15532.1"
            /db_xref="GI:2624257"
            /db_xref="SPTREMBL:O33280"
            /translation="MTVSCPSPSTSESEBOARLCLRLTARSRTAEALAGAKRGY
            PEDIGRVLDRLAAGLYDDTDFAPDMOYRAAKSKRALAEIHAQGVDDVIT
            VIGIDGAGERGRRAEKLVRRLREVLIIDGTGDEANYSRLVAMLRGCGTACEV
            VIELAEERRRRV"
            /complement(493..2865)
            /gene="recA"
            /complement(493..2865)
            /gene="recA"
            /note="RV2737c, (MTV002.02c), len: 790 aa. recA,
            identical to RECA_MCTU P26345 recA protein (790 aa).
            Contains self-splicing protein element (Intein) from 2294
            to 969 (c), similar to Intein II from TR:E332317
            (EMBL:Y13030) DNA-directed DNA polymerase (PC 2.7.7.7) from
            Thermococcus sp. (1829 aa), fasta scores: opt: 81 z-score:
            235.2 E(): 6e-06, 24.6% identity in 183 aa overlap.
            Contains PS00017 ATP/GTP-binding site motif A (P-loop),
            PS00321 recA signature, and PS00881 Protein splicing
            signature. See Davis et al. (1992) Cell 71(2):201-210"
            /codon_start=1
            /transl_table=11
            /product="recA"
            /protein_id="CA15533.1"
            /db_xref="GI:2624258"
            /db_xref="SWISS-PROT:P26345"
            /translation="MTQTPDRKALELAVNQIEKSYGKGSVMRLGDEARQISVIFPG
            SIADVALIGIGLGRGVIEIYESSG TVALHVAANAAGVAAAFIDADHADP
            DYAKLVGDTSDLSQPTGEALLETAL AIISGALDIIVDSVALVRALEEGM
            GDSHVIGQARLMGALKMGALNGSTGTAIFINLDRDIGMEGSEPTTGKALKE
            YASVMDYRREVTIKGTNAVGNTRKYKNCIACGTRIPDPVGTTHRIDGAG
            RKLHVVAAKAGDTLHARPVVSNPDGTQVYICIRAGCAIWAATPHKVLTEYGRRA
            AGELRKGDRAVQPRFDGDSAPTPADHARLGLYLIGDRDQWVGKTIININQR
            ALIDVVRIRIATLGCANPQGRISLAIAHRRGERNGVADLCQAGIYGLKWLKEETPN
            WFEPEPDIADIVGLTGLFESDGWSREQTGALRVGTTTSELAHQHMLLRGCV
            GSTVMDPQPKRSIVNGRIQSGROVFEVRIASGMDNTAAFEVPMQPPRAALIO
            AIPATGRRGSGATYLAEMTDVAVNYDERGVTNOEAAAMIGVAGSGRGMKVO
            LGASRLRRDRAVALADLDDKFLHDMLAEEELRSVIREVYPTRRARPDLEVEELTL
            VAEGVAVHNCSPPKQAEFDLITGKGISRESGLSDGVDGLIRKSGAMPTVEIGEOUG
            OGKEMANFLVENADVADETIKIKIKELIGAVVTDGDPNDGVLPAVDE"

```

```

misc_feature complement(790..2112)
              /gene="recA"
              /note="recA intein"
              complement(790..807)
              /gene="recA"
              /note="PS00881 protein splicing signature"
              complement(2197..2223)
              /gene="recA"
              /note="PS00321 recA signature"
              complement(2644..2667)
              /gene="recA"
              /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
              complement(2873..2877)
              /note="possible RBS upstream of RV2737c"
              complement(3247..3453)
              /gene="RV2738c"
              complement(3247..3453)
              /gene="RV2738c"
              /note="RV2738c, (MTV002.03c), len: 68 aa. Unknown,
              N-terminus is highly similar to the N-terminus of the
              upstream ORF MTV002.07c (78.4% identity in 37 aa overlap);
              also similar to A10209581SC4H8.5 Streptomyces coelicolor
              cosmid 4H8 (64aa) opt: 185 z-score: 283.5 E(): 2.9e-08;
              39.7% identity in 63 aa overlap"
              /codon_start=1
              /transl_table=11
              /product="hypothetical protein RV2738c"
              /protein_id="CA15534.1"
              /db_xref="GI:2624260"
              /db_xref="SPTREMBL:O33281"
              /translation="MLAGVRLTEFHERALHFGAAYGSVLDHVLGFGDSRAQAAT
              EDGVEPRDVRALCADFDPVDRW"
              complement(3464..4530)
              /gene="RV2739c"
              complement(3464..4530)
              /gene="RV2739c"
              /note="RV2739c, (MTV002.04c), len: 388 aa. Probable
              transferase, similar to eg. TRC 1560 (EMBL:L28170)
              pseudocaps aeruginosa thiamosyl transferase (426 aa)
              fasta scores: opt: 178 z-score: 226.3 E(): 1.9e-05, 25.9%
              identity in 425 aa overlap. Equivalent to Mycobacterium
              leprae protein MLCB33.02c (392 aa), fasta scores
              912947231MLCB33.2 Mycobacterium leprae cosmid B33 opt:
              2112 z-score: 2364.5 E(): 0; 80.9% identity in 388 aa
              overlap"
              /codon_start=1
              /transl_table=11
              /product="hypothetical protein RV2739c"
              /protein_id="CA15535.1"
              /db_xref="GI:2624261"
              /db_xref="SPTREMBL:O33282"
              /translation="KRVAVVAGPDDGHSFPAIALCQRRRAADPTLTGVELEEAR
              AAGIDAVEIDGLAATDRDIDAGARLHRRRAQAVANVRLALEDELVSQVITACGG
              MAELGIDPEWELNHPPLPSKGLPPIPSGALACTGRLGRARATMALGRSMRAG
              LRQRAAVREIELPARDPGLRLIATIPALEVPPDPAEAVVVGPLHFPDTRVLA
              IPAGTGPVVVAPSTALGTGTLGTVALPGLTGTGTVSGSRLVYSLGADLVPPM
              AVAGIGSQAELLIRADIVYGGGGMVAKTLLAGAPMVVVGGGGQMEIARVYRQGS
              AVILPRLTDAIVAAVNEVLSRPREARRARRAAVGAADPVRCHALALAG"
              4674..5123
              /gene="RV2740"
              4674..5123
              /gene="RV2740"
              /note="RV2740, (MTV002.05), len: 149 aa. unknown, similar
              to eg. leprae hypothetical protein TR:04950 (EMBL:294723)
              MLCB33.03 (82235_F3_140) (118 aa), fasta scores: opt: 498
              z-score: 538.9 E(): 7.3e-23, 51.6% identity in 161 aa
              overlap"
              /codon_start=1
              /transl_table=11
              /product="hypothetical protein RV2740"
              /protein_id="CA15536.1"
              /db_xref="GI:2624262"
              /db_xref="SPTREMBL:O33283"

```

QWVCGVEVDGRTILMRDYFDVYDMFKGLRLGLVALVPSKATL" /translation- "MAELTETSPETPETEATRAVEAFNLALONEDFDYDALGDLL VYENGFSRIRGRRATALLRMOGRVGFVKIHRIGADGAVALTERDALIIGPLRWG

gene

```

/ gene="PE_PGRS"
5355. 6932
CDS

```

```
/gene="PE_PGRS"  
/note="RV2141, (MTV002.06), len: 525 aa: Member of M  
tuberculosis PE_PGRS subfamily, similar to many eg.
```

Query Match	98.9%	Score 1221.4;	DB 1;	Length 5641;
Best Local Similarity	99.8%;	Pred. No. 1.7e-165;		
Matches 1233;	Conservative	0;	Mismatches 2;	Indels 1;
				Gaps 1

QY	1	ATCTTGACGATTAAATCGAACTTTCTTCACACTGGAAGCTACAGTATGAGAGGGGGTATC	60
Db	38201	ATCTTGACGATTAAATCGAACTTTCTTCACACTGGAAGCTACAGTATGAGAGGGGGTATC	38260
QY	61	ATGGCGCTGCGTATTTCGACCGAGACCAAAAACAACGAATTCGGGGTGGCCATACACCCG	120
Db	38261	ATGGCGCTGCGTATTTCGACCGAGACCAAAAACAAGATTCGGGGTGGCCATACACCCG	38320
QY	121	GCCGGCGTCGGGGAACCTAACCCGCTGTGGCCATGAGGTCATCCAGGACAGTGGCGGA	180
Db	38321	GCCGGCGTCGGGGAACCTAACCCGCTGTGGCCATGAGGTCATCCAGGACAGTGGCGGA	38380
QY	181	GAGGGCTGGGCTATCACCGACGCGGATTTCAAGGCGGACGCGCAACTGTGTGGCAC	240
Db	38381	GAGGGCTGGGCTATCACCGACGCGGATTTCAAGGCGGACGCGCAACTGTGTGGCAC	38440
QY	241	GCCGACACAGGTGTGGGGCCGACGCTATTTATTTGCTCAAGGTCAAAAGACGATAGCGCG	300
Db	38441	GCCGACACAGGTGTGGGGCCGACGCTATTTATTTGCTCAAGGTCAAAAGACGATAGCGCG	38500
QY	301	GAATACGCGCGCCCTCGACACAGGG - GATCTGTTCACGTTCTTGCAATTTGGCCGCTCA	359
Db	38501	GAATACGCGCGCCCTCGACACAGGGAGATCTTGTTCACGTTCTTGCAATTTGGCCGCTCA	38560
QY	360	CGTCTTGACACCGATGCGTTGTTCGATTTCGGGACACAGCTCAATTGCTTACGACCGTC	419
Db	38561	CGTCTTGACACCGATGCGTTGTTCGATTTCGGGACACAGCTCAATTGCTTACGACCGTC	38620
QY	420	CAGACCGCCGAGGGGGGCACTACCCGTGGCTGGCCGATGAGGAGATGGCGGTGACATC	479
Db	38621	CAGACCGCCGAGGGGGGCACTACCCGTGGCTGGCCGATGAGGAGATGGCGGTGACATC	38680
QY	480	GCGCGCCAGGTTGGCGCTTACCACTGATGCGAACCCAAAGGGGCGCGGTGTCTGATG	539
Db	38681	GCGCGCCAGGTTGGCGCTTACCACTGATGCGAACCCAAAGG - CCGGGTGTCTGATG	38740
QY	540	GGCGGGGTGTCGGGGGTGCAACCGGCGACGTCGTGTGATATGGGGCCGGACCGCGCGG	599
Db	38741	GGCGGGGTGTCGGGGGTGCAACCGGCGACGTCGTGTGATATGGGGCCGGACCGCGCGG	38800
QY	600	TACACGCGACCGCGCATCGCCCAAGGCAATGGGCGGACCGCTTACCGTTACATCAAC	659
Db	38801	TACACGCGACCGCGCATCGCCCAAGGCAATGGGCGGACCGCTTACCGTTACATCAAC	38860
QY	660	ATCGACAAACTTCGCACTCGACGCGGAGTTCTGGCGCGGATCCACTGCTGACTCA	719
Db	38861	ATCGACAAACTTCGCACTCGACGCGGAGTTCTGGCGCGGATCCACTGCTGACTCA	38920
QY	720	TGGGGCTAGAGAGTCGAGGGTGGCGGTCAAGCGTGGCCGCTGTGTTGGGCGCGCTG	779
Db	38921	TGGGGCTAGAGAGTCGAGGGTGGCGGTCAAGCGTGGCCGCTGTGTTGGGCGCGCTG	38980
QY	780	GTCGACAGCGGCCAGGCAACCCAAATTAGTCTGCAATTCATCTGTGGCGCATATGAACCA	839
Db	38981	GTCGACAGCGGCCAGGCAACCCAAATTAGTCTGCAATTCATCTGTGGCGCATATGAACCA	39040
QY	840	GGTGGGGTACTGGTGAATATAGCCATCGACACAGGGCGGCTGTTTCAGAGCTACGACCG	899
Db	39041	GGTGGGGTACTGGTGAATATAGCCATCGACACAGGGCGGCTGTTTCAGAGCTACGACCG	39100

QY	900	ACCACCTACGACCAACCCGAGCTTCGCGTGACGACGAGCGTTTAACTCGTCGGGGAC	959
Db	39101	ACCACCTACGACCAACCCGAGCTTCGCGTGACGACGAGCGTTTAACTCGTCGGGGAC	39160
QY	960	ATGCGCGCGCTCGGTGCGGAGACGCTGCACCTACGCGCTGACCAACGCGACGATGCCGTAT	1019
Db	39161	ATGCGCGCGCTCGGTGCGGAGACGCTGCACCTACGCGCTGACCAACGCGAGGATGCCGTAT	39220
QY	1020	GTGCTCGAGTTTCCGACCATGGCTGGCGGGGGGGCGTCCGCTGGAATCCGGCATAGGC	1079
Db	39221	GTGCTCGAGTTTCCGACCATGGCTGGCGGGGGGGCGTCCGCTGGAATCCGGCATAGGC	39280
QY	1080	AAAGGTCTTTCGACGACGACGAGGGCGCTACTCTGCCAACGGGTGGCCACCGACCTGGGG	1139
Db	39281	AAAGGTCTTTCGACGACGACGAGGGCGCTACTCTGCCAACGGGTGGCCACCGACCTGGGG	39340
QY	1140	GTGCGCTTACACGAGCGCCGACGCGTGTGGCGCTGACTCTGGCGCGCTCGTTACGCCGAN	1199
Db	39341	GTGCGCTTACACGAGCGCCGACGCGTGTGGCGCTGACTCTGGCGCGCTCGTTACGCCGAN	39400
QY	1200	CACACGTGGGAGTAGAGGGAACGAGATGTCGGCC	1235
Db	39401	CACACGTGGGAGTAGAGGGAACGAGATGTCGGCC	39436

**RESULT 12**

A87606	1235	h	DNA	PAT	22-JAN-2000
LOCUS	A87606				

LOCUS	Accession	DATE
DEFINITION	Sequence 3 from Patent WO9836089.	

ACCESSION	A87606
VERSION	A87606.1
	GI:6736246

**KEYWORDS** . unidentified source

SOURCE	unidentified
ORGANISM	unidentified

unclassified: 1 (bases 1 to 1000)

AUTHORS	Flohe, L. and Singh, M.
TEST KIT FOR TUBERCULOSIS	

Patent: WO 9836089-A 20-AUG-1998;

FEATURES  
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
Location/Qualifiers

```
source 1. .1235
        /organism="unidentified"
```

/db\_xref="taxon:32644"

BASE COUNT	236 a	394 c	382 g	420 t	3 Oct
ORIGIN					

Query Match	Score	DB 5;	Length
98.78;	1219;		

Best Local Similarity 99.6%; Pred. NO. 9.2e-165;  
Matches 1230; Conservative 0; Mismatches 4; Inde

1 ATCTTCAGATTAAATCGAATTTCTTTCACACTGAAGGTACAGTATCGA

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

DB 1 AATCTGCAAGTAATCGAAGCTTCTTCATCACACGGAAGGCAACAGTAACG

61 ATGGCGTGGTATTCGACCGAGACCAAAAACAAGAAATTCGGGGTGC

Db 61 ATGCGCGTCGGTATCCGACCGAGACCAAAACAACGAATCCGGGTGG

QY 121 GCCGCGTCGGGAACTAACCCGTCGTGGCAATGAGTGCTCATCCAGG

Bb 121 000000000000A#AA000000000000A#000000000000

181 GAGGGCTCGGCTATCACCGACGCGGATTTC AAGCGCGGCAGCGCGCAAC

[illegible]

101 GAGGGCCCGGCAICACCGACGCGGAIICAGGGGCGAGGCGCGCGCAAA

241 GCCGACCAGGTGTGGCCGACGCTGATTATTGCTCAAGGTCAAGAAG

Db 241 GCCGACCAGGTGTGGCCGACGCTGATTATTGCTCAAGGTCAGAAGAA



Db 781 GTGCCAGGCGCCAAAGGACCCAAATTAGTCTGAAATTCATTGTCGGCATATGAAACCA 840

QY 840 GGTGCGGACTAGTGTATATAGCATCGACGAGGCGGCTGTTGCAAGGCTCAAGACCG 899

Db 841 GGTGCGGACTAGTGTATATAGCATCGACGAGGCGGCTGTTGCAAGGCTCAAGACCG 900

QY 900 ACCACCTACGACCAACCCGACGCTGCGGTGACGACAGCGTGTCTTACTGCGTGGCGAAC 959

Db 901 ACCACCTACGACCAACCCGACGCTGCGGTGACGACAGCGTGTCTTACTGCGTGGCGAAC 960

QY 960 ATGGCCCGCTGCGGTGCGCAACAGCTGCGACCTACGCGCTGACCAACGCGACATGCGGTAT 1019

Db 961 ATGGCCCGCTGCGGTGCGCAACAGCTGCGACCTACGCGCTGACCAACGCGACATGCGGTAT 1020

QY 1020 GTGCTCGAGCTTGGCGACCATGCTGCGGCGGCGGCTGCGGTGCAATCGGCACTACCC 1079

Db 1021 GTGCTCGAGCTTGGCGACCATGCTGCGGCGGCGGCTGCGGTGCAATCGGCACTACCC 1080

QY 1080 AAAGGCTTTTGGACGACGAAAGGCGGCTTACTGTCCGAAAGGCGTGGCCACCGACCTGGGG 1139

Db 1081 AAAGGCTTTTGGACGACGAAAGGCGGCTTACTGTCCGAAAGGCGTGGCCACCGACCTGGGG 1140

QY 1140 GTGCGCTTACCGAGCCCGGCGAGGCTGCTGGCTGACCTCTGGCGGCTCTTAGCGCCGAN 1199

Db 1141 GTGCGCTTACCGAGCCCGGCGAGGCTGCTGGCTGACCTCTGGCGGCTCTTAGCGCCGAN 1200

QY 1200 CACACGTGCGGAGTAAGGAGAGCATGATGTGCGC 1234

Db 1201 CACACGTGCGGAGTAAGGAGAGCATGATGTGCGC 1235

RESULT 14

LOCUS A87608 1228 bp DNA PAT 22-JAN-2000

DEFINITION Sequence 5 from Patent M09836089.

VERSION A87608.1 GI:6736248

KEYWORDS

SOURCE unidentified.

ORGANISM unidentified.

REFERENCE 1 (bases 1 to 1228)

AUTHORS Flohe, L. and Singh, M.

TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE

JOURNAL Patent: WO 9836089-A-20-AUG-1998;

FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)

FEATURES

source 1..1228

1..1228

/organism="unidentified"

/db\_xref="taxon:32644"

BASE COUNT 236 a 391 c 382 g 219 t

ORIGIN

Query Match 98.3%; Score 1213.4; DB 5; Length 1228;

Best Local Similarity 99.8%; Pred. No. 5.8e-164;

Matches 1225; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 ATCTTGAGATTATTCGAACTTCTTCACTGAAGCTACAGTATCGAAGGGGTATC 60

Db 1 ATCTTGAGATTATTCGAACTTCTTCACTGAAGCTACAGTATCGAAGGGGTATC 60

QY 61 ATGCGCGTGTATTCGACGACGACCAAAAAAAGAAATTCGCGGTGGCCATCACCGCG 120

Db 61 ATGCGCGTGTATTCGACGACGACCAAAAAAAGAAATTCGCGGTGGCCATCACCGCG 120

QY 121 GCGCGCGTGTATTCGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 180

Db 121 GCGCGCGTGTATTCGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 180

QY 181 GAGGCGTGTATTCGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 240

Db 181 GAGGCGTGTATTCGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 240

QY 241 GCCGACAGGTGTGGGCGAGCGGTATTTATGCTCAAGGCAAGAACCGATAGCGCG 300

Db 241 GCCGACAGGTGTGGGCGAGCGGTATTTATGCTCAAGGCAAGAACCGATAGCGCG 300

QY 301 GAATACGCGCGCTGCGACACGCGC-GATCTGTTCAGTTCCTTTCATTTGGCGGCTCA 359

Db 301 GAATACGCGCGCTGCGACACGCGC-GATCTGTTCAGTTCCTTTCATTTGGCGGCTCA 360

QY 360 GGTCTTTCACCGATGCTGTTGGATTCGGGACACAGTCGAATGCTTACGAGACCTGC 419

Db 361 GGTCTTTCACCGATGCTGTTGGATTCGGGACACAGTCGAATGCTTACGAGACCTGC 420

QY 420 CAGACCGCGCGAGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 479

Db 421 CAGACCGCGCGAGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 480

QY 480 GCGCGCGAGGTTGGCGCTTACCACTGATGCAACCCAAAGGCGCGGCTGTGTATG 539

Db 481 GCGCGCGAGGTTGGCGCTTACCACTGATGCAACCCAAAGGCGCGGCTGTGTATG 540

QY 540 GCGCGGTTGCCCGGCTGCAACGCGCGGACGCTGTGTATCGGCGCGGACCGCGCG 599

Db 541 GCGCGGTTGCCCGGCTGCAACGCGCGGACGCTGTGTATCGGCGCGGACCGCGCG 600

QY 600 TACAAAGCAGCGCGCATCGCCAAAGGCGGCGCGGCGGCGGCTTACGATCAAC 659

Db 601 TACAAAGCAGCGCGCATCGCCAAAGGCGGCGGCGGCGGCGGCTTACGATCAAC 660

QY 660 ATGCAAACTTTCGCACTGACGCGGAGTTCGCGGCGGAGATCCACATCTCTACTCA 719

Db 661 ATGCAAACTTTCGCACTGACGCGGAGTTCGCGGCGGAGATCCACATCTCTACTCA 720

QY 720 TCGGC TACGAGCTGAGGTTGCCGTAACGTCGCGACCTGTGTATGGGCGCTCTG 779

Db 721 TCGGC TACGAGCTGAGGTTGCCGTAACGTCGCGACCTGTGTATGGGCGCTCTG 780

QY 780 GTGCGAGGCGCAAGGACCAAAATTAGTCTGCAATTCATTTGCGGCTATGAAACCA 839

Db 781 GTGCGAGGCGCAAGGACCAAAATTAGTCTGCAATTCATTTGCGGCTATGAAACCA 840

QY 840 GGTGCGTACTGTGTATATAGCATGACGACGAGGCGGCTGTTGCAAGGCTCACGACG 899

Db 841 GGTGCGTACTGTGTATATAGCATGACGACGAGGCGGCTGTTGCAAGGCTCACGACG 900

QY 900 ACCACCTACGACCAACCCGACGCTGCGGTGACGACAGCGTGTCTTACTGCGGCGAAC 959

Db 901 ACCACCTACGACCAACCCGACGCTGCGGTGACGACAGCGTGTCTTACTGCGGCGAAC 960

QY 960 ATGGCCCGCTGCGGTGCGCAACAGCTGACCTAGCGCTGACCAACGAGAGATGCGGTAT 1019

Db 961 ATGGCCCGCTGCGGTGCGCAACAGCTGACCTAGCGCTGACCAACGAGAGATGCGGTAT 1020

QY 1020 GTGCTCGAGTTGGCGACCATGCTGCGGCGGCGGCTGCGGTGCAATCGGCACTACCC 1079

Db 1021 GTGCTCGAGTTGGCGACCATGCTGCGGCGGCGGCTGCGGTGCAATCGGCACTACCC 1080

QY 1080 AAAGGCTTTTGGACGACGAAAGGCGGCTTACTGTCCGAAAGGCGTGGCCACCGACCTGGGG 1139

Db 1081 AAAGGCTTTTGGACGACGAAAGGCGGCTTACTGTCCGAAAGGCGTGGCCACCGACCTGGGG 1140

QY 1140 GTGCGCTTACCGAGCCCGGCGAGGCTGCTGGCTGACCTCTGGCGGCTCTTAGCGCCGAN 1199

Db 1141 GTGCGCTTACCGAGCCCGGCGAGGCTGCTGGCTGACCTCTGGCGGCTCTTAGCGCCGAN 1200

QY 1200 CACACGTGCGGAGTAAGGAGAGCATGATG 1227

Db 1201 CACACGTGCGGAGTAAGGAGAGCATGATG 1228

RESULT 15

LOCUS A89748 1228 bp DNA PAT 22-JAN-2000

DEFINITION Sequence 5 from Patent WO9832862.  
 ACCESSION A89748  
 VERSION A89748.1 GI:6738282  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1228)  
 FLOHE, L. and SINGH, M.  
 TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
 JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE, LEOPOLD (DE); SINGH, MAHAVIR (DE)  
 FEATURES  
 source 1..1228  
 Location/Qualifiers  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 236 a 391 c 382 g 219 t  
 ORIGIN

Query Match 98.3%; Score 1213.4; DB 5; Length 1228;  
 Best Local Similarity 99.8%; Pred. No. 5.8e-164;  
 Matches 1225; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```

QY 1 ATCTTGAGATTATCGAAGCTTTCTTACACTGAAGCGTACAGTATCGAGGGGTATC 60
DB 1 ATCTTGAGATTATCGAAGCTTTCTTACACTGAAGCGTACAGTATCGAGGGGTATC 60
QY 61 ATGGGCGTGGTATTCGACCGAGACCAAAACAGAAATTCGGGGGTGCGATACCCCG 120
DB 61 ATGGGCGTGGTATTCGACCGAGACCAAAACAGAAATTCGGGGGTGCGATACCCCG 120
QY 121 GCGGCGCTCGGAACTACCCGCTGCTGCGCATGAGGTGCTCATCCAGGAGGTGCCGA 180
DB 121 GCGGCGCTCGGAACTACCCGCTGCTGCGCATGAGGTGCTCATCCAGGAGGTGCCGA 180
QY 181 GAGGCGTGGCTATCACCGACCGCGATTTCAAGGCGCGGCGCACTGGTGGCAAC 240
DB 181 GAGGCGTGGCTATCACCGACCGCGATTTCAAGGCGCGGCGCACTGGTGGCAAC 240
QY 241 GCGGACGAGGTGGGCGGCGAGCGGTATTTGCTCAAGGTCAAGAACGATAGCGCG 300
DB 241 GCGGACGAGGTGGGCGGCGAGCGGTATTTGCTCAAGGTCAAGAACGATAGCGCG 300
QY 301 GAATACGCGCGCTGCGACACAGGCG -GATCTTGTTCAGCTTCTTGCAATTGCGCGCTCA 359
DB 301 GAATACGCGCGCTGCGACACAGGCGAGATCTTGTTCAGCTTCTTGCAATTGCGCGCTCA 359
QY 360 CGTCTTGACCGATGGGTGTTGATTCGGGCAACGATGCTCAATGGCTACGAGACCGTC 419
DB 360 CGTCTTGACCGATGGGTGTTGATTCGGGCAACGATGCTCAATGGCTACGAGACCGTC 419
QY 420 CAGACCGCGACGCGGACGCTACCTGCTTGGCCGATGAGCGAAGTGGCGGTGCACTC 479
DB 420 CAGACCGCGACGCGGACGCTACCTGCTTGGCCGATGAGCGAAGTGGCGGTGCACTC 479
QY 480 GCGGCGCAGGTTGGCGCTTACCACTGATGGAACCAAGGGGCGCGGTGCTGATG 539
DB 480 GCGGCGCAGGTTGGCGCTTACCACTGATGGAACCAAGGGGCGCGGTGCTGATG 539
QY 540 GCGGCGGCTGCGGCGGCTGCAACCGCGGCGAGCTGCTGATGCGGCGCGCGCGCG 599
DB 540 GCGGCGGCTGCGGCGGCTGCAACCGCGGCGAGCTGCTGATGCGGCGCGCGCGCG 599
QY 541 GCGGCGGCTGCGGCGGCTGCAACCGCGGCGAGCTGCTGATGCGGCGCGCGCGCG 600
DB 541 GCGGCGGCTGCGGCGGCTGCAACCGCGGCGAGCTGCTGATGCGGCGCGCGCGCG 600
QY 600 TACAACGAGCGCGCATGCGCAAGCGCATGGGCGCGACCGTTAGGTTCTAGACATCAAC 659
DB 600 TACAACGAGCGCGCATGCGCAAGCGCATGGGCGCGACCGTTAGGTTCTAGACATCAAC 659
QY 660 ATCGACAAACTTGGGCAACTGACGCGGAGTTGGGCGCGGATCCACACTCGCTACTCA 719
DB 660 ATCGACAAACTTGGGCAACTGACGCGGAGTTGGGCGCGGATCCACACTCGCTACTCA 719
QY 720 ATCGACAAACTTGGGCAACTGACGCGGAGTTGGGCGCGGATCCACACTCGCTACTCA 720
DB 720 ATCGACAAACTTGGGCAACTGACGCGGAGTTGGGCGCGGATCCACACTCGCTACTCA 720
QY 720 TCGGCGTACGAGCTGAGGGTGGCGTCAAAAGTCCGACCTGCTGATTTGGGCGCTCTG 779

```

```

DB 721 TCGGCGTACGAGCTGAGGGTCCGCTCAAAAGTCCGACACTGCTGATGGGCGCTCTG 780
QY 780 GTGCCAGGCGCCAAAGGACCCCAATTAAGTCTGATTTACTGTTGCGCATATGAACCA 839
DB 781 GTGCCAGGCGCCAAAGGACCCCAATTAAGTCTGATTTACTGTTGCGCATATGAACCA 840
QY 840 GTGCCAGTCTGATATTAAGCATGACAGGCGCGGCTTTGGAAGGCTACAGACCG 899
DB 841 GTGCCAGTCTGATATTAAGCATGACAGGCGCGGCTTTGGAAGGCTACAGACCG 900
QY 900 ACCACTACGACCCAGCGCTTCCGCGTGACGACAGCGCTGTTTACTGCGTGGCGAAC 959
DB 901 ACCACTACGACCCAGCGCTTCCGCGTGACGACAGCGCTGTTTACTGCGTGGCGAAC 960
QY 960 ATGCCCGCTCGTGGCGGAGAGCGTACGACTACGCGGTGACCAAGCGAGATGCGCTAT 1019
DB 961 ATGCCCGCTCGTGGCGGAGAGCGTACGACTACGCGGTGACCAAGCGAGATGCGCTAT 1020
QY 1020 GTGCTGAGCTTGGCGACCATGAGTGGCGGCGGCTGCGGCTGCAATCCGCACTAGCC 1079
DB 1021 GTGCTGAGCTTGGCGACCATGAGTGGCGGCGGCTGCGGCTGCAATCCGCACTAGCC 1080
QY 1080 AAAGGCTTTTCAAGCAGCAAGGCGGCTTACTGTCGAAGCGGTGGCCACCACTGGGG 1139
DB 1081 AAAGGCTTTTCAAGCAGCAAGGCGGCTTACTGTCGAAGCGGTGGCCACCACTGGGG 1140
QY 1140 GTGCGTTACGAGGCGCGCCAGCGCTGCTGCGCTGACTCTGCGCGCTGTTACCGCGAN 1199
DB 1141 GTGCGTTACGAGGCGCGCCAGCGCTGCTGCGCTGACTCTGCGCGCTGTTACCGCGAN 1200
QY 1200 CACACGTCGGAGTAAGGAAGCGATGA 1227
DB 1201 CACACGTCGGAGTAAGGAAGCGATGA 1228

```

Search completed: June 22, 2000, 14:54:29  
 Job time: 17684 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:08:21; Search time 458.59 Seconds  
(without alignments)  
673.777 Million cell updates/sec

Title: US-09-362-485-6

Perfect score: 1235  
Sequence: 1 ATCTTGCGAGATTATCGAAC.....GGCAGCGATGATGCGGCC 1235

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1207	97.7	1245	1 V49626	Mycobacterium tube
2	1207	97.7	1260	1 V49510	Mycobacterium sp.
3	1207	97.7	1260	1 V49625	Mycobacterium tube
4	455.6	36.9	682	1 V49511	Mycobacterium mari
5	266.8	21.6	1125	1 N91423	Sequence of heat-r
6	202	16.4	28171	1 V52155	Streptococcus pneu
7	168	13.6	9280	1 W74442	Streptococcus pneu
8	87.6	7.1	1074	1 T67971	H. pylori membrane
9	69.2	5.6	544	1 V43039	Streptococcus pneu
10	56.8	4.6	31096	1 V74370	Staphylococcus aur
11	50.8	4.1	390	1 O21833	Randomising oligon
12	50.8	4.1	390	1 O36859	PCR primer for 5'
13	50.4	4.1	15872	1 T68715	Streptomyces venez
14	46.4	3.8	985	1 V44439	Mycobacterium tube
15	46.4	3.8	985	1 V44548	M. tuberculosis im
16	46	3.7	2582	1 T73117	Actinoplanes sp. a
17	46	3.7	17955	1 V56642	Actinoplanes sp. a
18	45.4	3.7	833	1 O64203	snab gene encoding
19	45.4	3.7	5392	1 O64201	Sequence compriain
20	45	3.6	3946	1 T93610	Mycobacterium tube
21	44.8	3.6	535	1 V44428	Mycobacterium tube
22	44.8	3.6	535	1 V44537	M. tuberculosis im
23	44.8	3.6	11485	1 X53481	Human adenosine Al
24	43.6	3.5	20856	1 T66750	Actinoplanes sp. a
25	43.6	3.5	28958	1 T89556	Actinoplanes sp. a
26	43.6	3.5	49377	1 V05287	The soraphen biosy
27	42.4	3.4	1833	1 O64206	sbR gene encoding
28	42.4	3.4	2185	1 V84066	Clone p5-4 encodin
29	42.4	3.4	2185	1 O99365	S. lividans protea
30	42.4	3.4	24379	1 T93095	Streptomyces feno
31	42.4	3.4	24379	1 V25925	Streptomyces roseo
32	42.2	3.4	329	1 V44425	Mycobacterium tube
33	42.2	3.4	329	1 V44534	M. tuberculosis im
34	42	3.4	3946	1 T93610	Mycobacterium tube

C	35	42	3.4	4018	1	Q63879	Polyhydroxyalkanoa
	36	41.8	3.4	882	1	V44403	Mycobacterium tube
	37	41.8	3.4	882	1	V64512	M. tuberculosis im
	38	41.6	3.4	1413	1	X34205	Mycobacterium spec
	39	41.6	3.4	1722	1	X34206	Mycobacterium spec
	40	41.6	3.4	12036	1	Q04668	FHA structural gen
	41	41.2	3.3	2414	1	Q05926	Sequence encoding
	42	41	3.3	29879	1	Q46806	eryA region of S.
C	43	40.8	3.3	2151	1	Q76252	HSV-2 protease, IC
C	44	40.8	3.3	2151	1	Q75251	HSV-2 protease/ICP
C	45	40.8	3.3	2472	1	Q84671	HSV-2 UL26 gene. N

## ALIGNMENTS

RESULT	1	
V49626		
ID	V49626 standard; DNA; 1245 BP.	
AC	V49626;	
DT	20-NOV-1998 (first entry)	
DE	Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.	
KM	ss: Alanine dehydrogenase: tuberculosis; TB; mycobacteria; L-alanine.	
OS	Mycobacterium tuberculosis.	
PN	W0836089-A2.	
PD	20-AUG-1998.	
PF	29-JAN-1998; E00483.	
PR	29-JAN-1997; EP-101338.	
PA	(FLOH/) FLOHE L.	
PI	Flohe L, Hutter B, Kolk A, Singh M;	
DR	WPI: 98-457123/39.	
PT	Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity	
PT	- useful for, e.g. for diagnosis, differentiation of strains,	
PT	monitoring vaccination and identification of mycobacterial	
PT	inhibitors	
PS	Disclosure; Fig 3.19; 5'p; German.	
CC	The Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in	
CC	the production of kits for diagnosing tuberculosis (TB) and other	
CC	mycobacterial infections in humans or animals. Kits are used for direct	
CC	diagnosis of TB on clinical samples (e.g. body fluids) and can	
CC	differentiate between pathogenic and non-virulent strains, e.g. for	
CC	identifying pathogens of the M. tuberculosis (M. t.) complex. The kit may	
CC	also be used to identify substances that inhibit mycobacteria, for	
CC	combating epidemics and for vaccination follow-up. Oligonucleotides	
CC	derived from AlaDH are used for sim. lab. in diagnostic hybridisation tests,	
CC	also for culture confirmation of isolated strains and for chromosome	
CC	fingerprinting to detect/differentiate between mycobacteria, and for	
CC	L-alanine-specific biotransformation reactions. AlaDH is an early	
CC	antigen, present extracellularly after only a few days of growth, making	
CC	it an ideal drug target.	
SQ	Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;	

Query Match 97.7%; Score 1207; DB 1; Length 1245;  
Best Local Similarity 99.4%; Pred. No. 3.7e-242;  
Matches 1234; Conservative 0; Mismatches 1; Indels 7; Gaps 2;

QY	1	ATCTTGCGAGATTATCGAACTTTCTTCACACTGAAGGCTACAGATGAGGGGTATC	60
DB	1	ATCTTGCGAGATTATCGAACTTTCTTCACACTGAAGGCTACAGATGAGGGGTATC	60
QY	61	ATGGCGGTGGTATTCGACCGAGACCAAAACAGC-----AATTCGGGTGGCCATC	114
DB	61	ATGGCGGTGGTATTCGACCGAGACCAAAACAGC-----AATTCGGGTGGCCATC	120
QY	115	ACCCGCGCGCGGTGGCGGAGTAAACCGTGTGGCATGAGTGCATCAGCAGGT	174
DB	121	ACCCGCGCGCGGTGGCGGAGTAAACCGTGTGGCATGAGTGCATCAGCAGGT	180
QY	175	GCCGAGAGGGGTGGCGTATCAGCGAGCGGATTTCAAGCGGCGGCGCAACTGTC	234
DB	181	GCCGAGAGGGGTGGCGTATCAGCGAGCGGATTTCAAGCGGCGGCGCAACTGTC	240



```

QY 235 GGCACCCGCGACAGGTGTGGGCGACGCTGATTTATGCTCAAGGTCAAGAACCGATA 294
    |||
DB 241 GGCACCCGCGACAGGTGTGGGCGACGCTGATTTATGCTCAAGGTCAAGAACCGATA 300
QY 295 GCGGCGGAATAGCGCGGCTCGGACACGGGC-GATCTTGTCTAGCTTCTGCAATTTGGC 353
    |||
DB 301 GCGGCGGAATAGCGCGGCTCGGACACGGGC-GATCTTGTCTAGCTTCTGCAATTTGGC 360
QY 354 GCGTCACGTGCTTGCACGATGCTTGTGTGATTCGCGACACGCTCAATTGCTTACGAG 413
    |||
DB 361 GCGTCACGTGCTTGCACGATGCTTGTGTGATTCGCGACACGCTCAATTGCTTACGAG 420
QY 414 ACCGTCCAGACCGCGGACGCGGACCTACCTCTGTCGCCGATGAGCGAAATGCGCGGT 473
    |||
DB 421 ACCGTCCAGACCGCGGACGCGGACCTACCTCTGTCGCCGATGAGCGAAATGCGCGGT 480
QY 474 CGATCCGCGCGCGAGGTGTGGGCTTACACCTGATGCGAACCCAAAGGGGGCGGGGTGTG 533
    |||
DB 481 CGATCCGCGCGCGAGGTGTGGGCTTACACCTGATGCGAACCCAAAGGGGGCGGGGTGTG 540
QY 534 CTGATGGGCGGGGTGCGCGGCTGGAACCGGCGCAGCTG-GGTGATCGGGCGCGGACCC 593
    |||
DB 541 CTGATGGGCGGGGTGCGCGGCTGGAACCGGCGCAGCTG-GGTGATCGGGCGCGGACCC 600
QY 594 GCGGCGTACAAAGGACGCGCGCATGCGCAAGCGCATGGGCGGACGCTTACGCTTACGAC 653
    |||
DB 601 GCGGCGTACAAAGGACGCGCGCATGCGCAAGCGCATGGGCGGACGCTTACGCTTACGAC 660
QY 654 ATCAACATCGACAAATTTGCGGAACCTGACGCGCGAGTTCTGGGCGCGGATCCACACTGC 713
    |||
DB 661 ATCAACATCGACAAATTTGCGGAACCTGACGCGCGAGTTCTGGGCGCGGATCCACACTGC 720
QY 714 TACATATCGGCTGACGAGCTGAGAGGTGCGGTCAAAAGTGGCGGACCTGAGTATGGGGCC 773
    |||
DB 721 TACATATCGGCTGACGAGCTGAGAGGTGCGGTCAAAAGTGGCGGACCTGAGTATGGGGCC 780
QY 774 GTCTGTGTGCGAGGCGCGCAAGGACCCAAATTAATGCTGCAATTCATCTGTGCGCATATG 833
    |||
DB 781 GTCTGTGTGCGAGGCGCGCAAGGACCCAAATTAATGCTGCAATTCATCTGTGCGCATATG 840
QY 834 AAACCAAGTGGGTATGATGATGATATGACCATGACCAAGGGCGGCTTTTCGAAGGCTCA 893
    |||
DB 841 AAACCAAGTGGGTATGATGATGATATGACCATGACCAAGGGCGGCTTTTCGAAGGCTCA 900
QY 894 CGACGACGACACCTACGACCAACCCGACGTTGCGCGTGCACGACGACGCTGTTTACTGGGTG 953
    |||
DB 901 CGACGACGACACCTACGACCAACCCGACGTTGCGCGTGCACGACGACGCTGTTTACTGGGTG 960
QY 954 GCGAATATGCGCGCTGTGGTGGCGAAGACGTGACCTTACGCGCTGACCAACGCGAGATG 1013
    |||
DB 961 GCGAATATGCGCGCTGTGGTGGCGAAGACGTGACCTTACGCGCTGACCAACGCGAGATG 1020
QY 1014 CCGATATGCTGAGGCTTGGCGACCATGGCTGGGGGGGGGGCGGTGGATTCGGCA 1073
    |||
DB 1021 CCGATATGCTGAGGCTTGGCGACCATGGCTGGGGGGGGGGCGGTGGATTCGGCA 1080
QY 1074 CTAGCCAAAGGTCTTTTGCAGCGACGAAGGGGCGTTACTGTTCGGAACGGGTGGCCACCGAC 1133
    |||
DB 1081 CTAGCCAAAGGTCTTTTGCAGCGACGAAGGGGCGTTACTGTTCGGAACGGGTGGCCACCGAC 1140
QY 1134 CTGGGGGTGCGCTTACCGAGCGCGCGGACGAGGTGTGCTGCTGAGCTCTTGGCGGCTGTTAC 1193
    |||
DB 1141 CTGGGGGTGCGCTTACCGAGCGCGCGGACGAGGTGTGCTGCTGAGCTCTTGGCGGCTGTTAC 1200
QY 1194 GCGGACGACAGCTCGGGAGTAAAGGAAGCATGATGTCGGCC 1235
    |||
DB 1201 GCGGACGACAGCTCGGGAGTAAAGGAAGCATGATGTCGGCC 1242

```

RESULT 2  
V49510  
ID V49510 standard; DNA; 1260 BP.  
AC V49510;

20-OCT-1998 (first entry)  
Mycobacterium sp. Aladh DNA.  
KM Alanine dehydrogenase; Aladh; ADH; diagnosis; tuberculosis; pathogen;  
KM swimmers disease; vaccine; epidemic; infection; identification; ss.  
OS Mycobacterium sp.  
PN M09832862-A2.  
PF 30-JUL-1998.  
PR 29-JAN-1998: E00484.  
PR 29-JAN-1997: EP-101339.  
PA (FLOH/) FLOHE L.  
PI Flohe L, Hutter B, Kolk A, Singh M.  
DR MPI: 98-427958/36.  
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum  
PT - used for diagnosis of tuberculosis and other mycobacterial  
PT diseases, also for treatment and prevention, for drug screening and  
PT for bio-transformation  
PS disclosure, page 11; 57pp; German.  
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated  
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis  
CC and other mycobacterial infections (including 'swimmers disease', caused  
CC by M. marinum, a fish pathogen) in humans or animals. The protein can  
CC also be used for control of epidemics and for vaccination, to screen for  
CC agents with anti-mycobacterial activity, and in bio-transformations that  
CC are specific for L-alanine. Also mycobacteria can be identified by  
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted  
CC early during infection.  
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 97.7%; Score 1207; DB 1; Length 1260;  
Best local similarity 99.4%; Pred. No. 3.7e-242;  
Matches 1234; Conservative 0; Mismatches 1; Indels 7; Gaps 2;

```

QY 1 ATCTTCAGATTAATGCACTTCTTACACTGAAAGCGTACAGTATCGAGAGGGTAAATC 60
    |||
DB 16 ATCTTCAGATTAATGCACTTCTTACACTGAAAGCGTACAGTATCGAGAGGGTAAATC 75
QY 61 ATCCGCGTGGATATCCGA :GAGACAAAACAAAG-----AATCCGGGTGGCATC 114
    |||
DB 76 ATCCGCGTGGATATCCGA :GAGACAAAACAAAG-----AATCCGGGTGGCATC 135
QY 115 ACCCGCGCGGCGTGGCGGAACTACCCGCTGGGCATGAGTGGCTATCCAGGCAAGT 174
    |||
DB 136 ACCCGCGCGGCGTGGCGGAACTACCCGCTGGGCATGAGTGGCTATCCAGGCAAGT 195
QY 175 GCGGAGAGGGCTCGGATATCACACGCGGATTTCAAGGCGGACGCGCACTGCTC 234
    |||
DB 196 GCGGAGAGGGCTCGGATATCACACGCGGATTTCAAGGCGGACGCGCACTGCTC 255
QY 235 GCGACGCGGACAGGTGGGGCGGACGCTGATTTATGCTCAAGGTCAAGAACCGATA 294
    |||
DB 256 GCGACGCGGACAGGTGGGGCGGACGCTGATTTATGCTCAAGGTCAAGAACCGATA 315
QY 295 GCGGCGGAATAGCGCGGCTCGGACACGGGC-GATCTTGTCTAGCTTCTGCAATTTGGC 353
    |||
DB 316 GCGGCGGAATAGCGCGGCTCGGACACGGGC-GATCTTGTCTAGCTTCTGCAATTTGGC 375
QY 354 GCGTCACGTGCTTGCACGATGCTTGTGATTCGCGACACGATCAATTGCTTACGAG 413
    |||
DB 376 GCGTCACGTGCTTGCACGATGCTTGTGATTCGCGACACGATCAATTGCTTACGAG 435
QY 414 ACCGTCCAGACCGCGGACGCGGACCTACCTCTGTCGCCGATGAGCGAAATGCGCGGT 473
    |||
DB 436 ACCGTCCAGACCGCGGACGCGGACCTACCTCTGTCGCCGATGAGCGAAATGCGCGGT 495
QY 474 CGACTGCGCGCGAGGTGGCGCTTACCACTGATGCGAAACCAAGGGGGCGGGGTGTG 533
    |||
DB 496 CGACTGCGCGCGAGGTGGCGCTTACCACTGATGCGAAACCAAGGGGGCGGGGTGTG 555
QY 534 CTGATGGGCGGGGTGCGCGGCTGGAACCGGCGGACGCTGATGATGCGGGCGGACCC 593
    |||
DB 556 CTGATGGGCGGGGTGCGCGGCTGGAACCGGCGGACGCTGATGATGCGGGCGGACCC 615

```

QY 594 GCGGCTACAAAGGAGCCGATGCGCAAGCGCATGGGCGCGAGCCGTTAGCGTTAGAC 653  
DB 616 GCGGCTACAAAGGAGCCGATGCGCAAGCGCATGGGCGCGAGCCGTTAGCGTTAGAC 675  
QY 654 ATCAACATCGCAAACTTGGCACTGACGCGCGAGTTCTGGCGCGGATCCACATCGC 713  
DB 676 ATCAACATCGCAAACTTGGCACTGACGCGCGAGTTCTGGCGCGGATCCACATCGC 735  
QY 714 TACTCATCGGCTACGAGCTGAGAGGTGCGCTAAAGCTGCGCACTGTGATTTGGGCG 773  
DB 736 TACTCATCGGCTACGAGCTGAGAGGTGCGCTAAAGCTGCGCACTGTGATTTGGGCG 795  
QY 774 GTCTGTGTCAGGCGCGCAAGGACCAAAATTAATTAATTAATTAATTAATTAATTA 833  
DB 796 GTCTGTGTCAGGCGCGCAAGGACCAAAATTAATTAATTAATTAATTAATTAATTA 855  
QY 834 AAACCAAGTGGGCTAGTGTGATATAGCATTCAGCAAGGCGGCTTTTGAAGGCTCA 893  
DB 856 AAACCAAGTGGGCTAGTGTGATATAGCATTCAGCAAGGCGGCTTTTGAAGGCTCA 915  
QY 894 CGACCGACCACTACGACACCGCGAGCTTGGCGTGCAGCAGCGCTTTTACTGCGTG 953  
DB 916 CGACCGACCACTACGACACCGCGAGCTTGGCGTGCAGCAGCGCTTTTACTGCGTG 975  
QY 954 GCGAACAATGCGGCTGGGTGCGCAAGCGTACGACGCGTACGCAACGCGACGATG 1013  
DB 976 GCGAACAATGCGGCTGGGTGCGCAAGCGTACGACGCGTACGCAACGCGACGATG 1035  
QY 1014 CCGATATGCTGAGCTTGGCGACCATATGCTGGCGGCGGCGTCCGCTGCAATCGGCA 1073  
DB 1036 CCGATATGCTGAGCTTGGCGACCATATGCTGGCGGCGGCGTCCGCTGCAATCGGCA 1095  
QY 1074 CTAGCCAAAGGTCTTTGAGCGACGAGAGGGCGTTACTGTCCGAAAGGGTGGCCACGAC 1133  
DB 1096 CTAGCCAAAGGTCTTTGAGCGACGAGAGGGCGTTACTGTCCGAAAGGGTGGCCACGAC 1155  
QY 1134 CTGGGGGGCGGCTACGAGCGCGCGAGCGTGTGCTGACCTGGCGGCTGCTTAC 1193  
DB 1156 CTGGGGGGCGGCTACGAGCGCGCGAGCGTGTGCTGACCTGGCGGCTGCTTAC 1215  
QY 1194 GCGGCAACAGCTGCGGAGTAAAGGAAAGCGATGATGCGGC 1235  
DB 1216 GCGGCAACAGCTGCGGAGTAAAGGAAAGCGATGATGCGGC 1257

RESULT 3  
V49625  
ID V49625 standard: DNA; 1260 BP.  
AC V49625:  
DT 20-NOV-1998 (first entry)  
DE Mycobacterium tuberculosis Alanae dehydrogenase.  
KW ss: Alanae dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.  
OS Mycobacterium tuberculosis.  
PN M09836089-A2.  
PD 20-AUG-1998  
PF 20-JAN-1998: E00483  
PR 20-JAN-1997: EP-101338.  
PA (FLOH/), FLOH L.  
PI Flohe L, Hutter B, Kolk A, Singh M;  
DR WPI; 98-457123/39.  
PT Diagnosis of, e.g. tuberculosis from alanae dehydrogenase activity  
PT - useful for, e.g. for diagnosis, differentiation of strains,  
PT monitoring vaccination and identification of mycobacterial  
PT inhibitors  
PS Claim 13; Fig 2.3; 55pp; German.  
CC The Mycobacterium tuberculosis alanae dehydrogenase (AlaDH) is in  
CC the production of kits for diagnosing tuberculosis (TB) and other  
CC mycobacterial infections in humans or animals. Kits are used for direct  
CC diagnosis of TB on clinical samples (e.g. body fluids) and can  
CC differentiate between pathogenic and non-virulent strains, e.g. for  
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may  
CC also be used to identify substances that inhibit mycobacteria, for  
CC combatting epidemics and for vaccination follow-up. Oligonucleotides

CC derived from AlaDH are used similarly in diagnostic hybridisation tests,  
CC also for culture confirmation of isolated strains and for chromosome  
CC fingerprinting to detect/differentiate between mycobacteria, and for  
CC L-alanine-specific biotransformation reactions. AlaDH is an early  
CC antigen, present extracellularly after only a few days of growth, making  
CC it an ideal drug target.  
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;  
Query Match 97.7%; Score 1207; DB 1; Length 1260;  
Best Local Similarity 99.4%; Pred. No. 3.7e-242;  
Matches 1234; Conservative 0; Mismatches 1; Indels 7; Gaps 2;  
QY 1 ATCTTGAGATTAATCGAATTTCTTACACTGAAGGCTACGATGAGAGGGATATC 60  
DB 16 ATCTTGAGATTAATCGAATTTCTTACACTGAAGGCTACGATGAGAGGGATATC 75  
QY 61 ATGCGGCTGGTATTCGACCGAGACCAAAACACG-----AATTCCGGTGGCCATC 114  
DB 76 ATGCGGCTGGTATTCGACCGAGACCAAAACACGAAATTCATTCGGGTGGCCATC 135  
QY 115 ACCCGGCGCGCGGCGGAGAACTAACCCGCTGGCCATGAGGCTCATCCAGGAGGT 174  
DB 136 ACCCGGCGCGCGGCGGAGAACTAACCCGCTGGCCATGAGGCTCATCCAGGAGGT 195  
QY 175 GCGGAGAGGGCTCGGCTATCACCGACGCGATTTCAAGCGCGGCGGCGCAACTGTC 224  
DB 196 GCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGCGCGGCGGCGCAACTGTC 255  
QY 235 GGCACCGCCGACAGGTGTGGCGGAGCGTGTATTTATGCTCAAGGTCAAAGAACGATA 294  
DB 256 GGCACCGCCGACAGGTGTGGCGGAGCGTGTATTTATGCTCAAGGTCAAAGAACGATA 315  
QY 295 GCGGAGAAATAGGCGCGCTGCGACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 353  
DB 316 GCGGAGAAATAGGCGCGCTGCGACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 375  
QY 354 GCGTACAGTCTTTCACCGATGCGTGTGATTCGCGGACCAAGCTAATTCCTACGAG 413  
DB 376 GCGTACAGTCTTTCACCGATGCGTGTGATTCGCGGACCAAGCTAATTCCTACGAG 435  
QY 414 ACCGTCAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 473  
DB 436 ACCGTCAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 495  
QY 474 CGACTGCGCGCGCGGCTGGCGGCTTACCATCATGTCGAACCGCAAGGGGCGGCGGCTG 533  
DB 496 CGACTGCGCGCGCGGCTGGCGGCTTACCATCATGTCGAACCGCAAGGGGCGGCGGCTG 555  
QY 534 CTGATGGGCGGGGTGCGGCGGCTGGAACCGCGGAGCTGCTGATGCGGCGGCGGCGG 593  
DB 556 CTGATGGGCGGGGTGCGGCGGCTGGAACCGCGGAGCTGCTGATGCGGCGGCGGCGG 615  
QY 594 GCGGCTACAAAGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 653  
DB 616 GCGGCTACAAAGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 675  
QY 654 ATCAACATCGCAAACTTGGCAACTGACGCGCGAGTTCTGGCGCGGATCCACATCGC 713  
DB 676 ATCAACATCGCAAACTTGGCAACTGACGCGCGAGTTCTGGCGCGGATCCACATCGC 735  
QY 714 TACTCATCGGCTACGAGCTGAGAGGTGCGCTAAAGCTGCGCACTGTGATTTGGGCG 773  
DB 736 TACTCATCGGCTACGAGCTGAGAGGTGCGCTAAAGCTGCGCACTGTGATTTGGGCG 795  
QY 774 GTCTGTGTCAGGCGCGCAAGGACCAAAATTAATTAATTAATTAATTAATTAATTA 833  
DB 796 GTCTGTGTCAGGCGCGCAAGGACCAAAATTAATTAATTAATTAATTAATTAATTA 855  
QY 834 AAACCAAGTGGGCTAGTGTGATATAGCATTCAGCAAGGCGGCTTTTGAAGGCTCA 893  
DB 856 AAACCAAGTGGGCTAGTGTGATATAGCATTCAGCAAGGCGGCTTTTGAAGGCTCA 915

QY	894	CGACGACACACACTACAGACACACCCGAGCTTGCCGCTACAGAACACCTGTTTACAGCTG	953
Db	916	CGACCGACACACTACAGACACACCCGAGCTTGCCGCTACAGAACACCTGTTTACAGCTG	975
QY	954	GGCAACATGCCCCCTCTGGTGGCCGAAGACGTGCACCTACACCGCTGCACACGACGATG	1013
Db	976	GGCAACATGCCCCCTCTGGTGGCCGAAGACGTGCACCTACACCGCTGCACACCGACGATG	1035
QY	1014	CCGTAATGCTCGAGCTTGGCCGACCAATGCTGGCGGGCGCGTGGCGGTGCAGATCCGACA	1073
Db	1036	CCGTAATGCTCGAGCTTGGCCGACCAATGCTGGCGGGCGCGTGGCGGTGCAGATCCGACA	1095
QY	1074	CTAGCCAAAGGCTCTTTTGAGAGCAGCAAGGGGGCTTACTCCGAAGGGTGGCCACACGAC	1133
Db	1096	CTAGCCAAAGGCTCTTTTGAGAGCAGCAAGGGGGCTTACTCCGAAGGGTGGCCACACGAC	1155
QY	1134	CTGGGGGTGCGCTTCACCGAGCCCGCCAGCGTCTGAGCTCTCGGCGCGCTGTTAC	1193
Db	1156	CTGGGGGTGCGCTTCACCGAGCCCGCCAGCGTCTGAGCTCTCGGCGCGCTGTTAC	1215
QY	1194	GGCGAACAACAGCTGGGGAGTAAAGGAAGCATATGCGGC	1215
Db	1216	GGCGAACAACAGCTGGGGAGTAAAGGAAGCATATGCGGC	1257

```

RESULT      4
V49511
ID      V49511 standard; DNA; 682 BP.
AC      V49511:
DT      20-OCT-1998 (first entry)
DE      Mycobacterium marinum Mar3 DNA.
KW      Alanine dehydrogenase; ADH; ADH; diagnosis; tuberculosis; pathogen;
        swimmers disease; vaccine; epidemic; infection; identification; ss.
OS      Mycobacterium marinum.
FH      Key
FT      Location/Qualifiers
FT      1..682
FT      /*tag=a
FT      /codon_start=3
FT      /product="Mar3"
FT      /note="Alanine dehydrogenase"
PN      MO9832862-A2.
PD      30-JUL-1998.
PF      29-JAN-1998; E00484.
PR      29-JAN-1997; EP-101339.
PA      (FLOH/) FLOHE L.
PI      Flohe L, Hutter B, Kolx A, Singh M;
DR      MPI; 98-427958/36.
P-PSDB: M64481.
NC      Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT      used for diagnosis of tuberculosis and other mycobacterial
PT      diseases, also for treatment and prevention, for drug screening and
PT      for bio-transformation
PS      Claim 1: Page 34-35; 57pp; German.
CC      This sequence encodes an alanine dehydrogenase (ADH) protein, Mar3
CC      isolated from Mycobacterium marinum. This protein is used to diagnose
CC      tuberculosis and other mycobacterial infections (including 'swimmers'
CC      disease', caused by M. marinum, a fish pathogen) in humans or animals.
CC      The protein can also be used for control of epidemics and for
CC      vaccination, to screen for agents with anti-mycobacterial activity, and
CC      bio-transformations that are specific for L-alanine. Also mycobacteria
CC      can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC      that is secreted early during infection.
SQ      Sequence 682 BP; 105 A; 254 C; 225 G; 98 T;

```

OY	158	TGCTCATCCAGGACAGGTGGCCGAGAAAGGGGCTTCGGCTATACACGAGACGGCGGATTTCAAGGCGG	217
Db	61	TGCTGATCCAGGCGGGTGGCGGAGAAAGGCTTCGGCATTCGCGAGCGGCACTTCAAGGCGG	120
OY	218	CAGGCGCGCAACTGCTGGGCGCACCCGCCACCAAGTGTGTGGGCCGAGCGCTGATTTATTTGCTCA	277
Db	121	CCGGTGGCCACTGATATACGACCGCCGCCACAGGTGTGGGGGAGATGGCCGACTCTGCTCA	180
OY	278	AGGTCAAGAAGCCGATAGCGCGGAGAAATACGAGCCGCTCTGCACACAGGGGCGC-GATGCTTTGTCA	336
Db	181	AGGTCAAGAAGCCGATGAGTCCGAGTACGGCCGGCTTGGCGCGGGGCGCAACACCTGTGTCA	240
OY	337	CGTTCTTGCACTTTTGGCGCGGTACAGTCTGTGCACCGATGGCGTGTGTGGATTCGGGGACCA	396
Db	241	CGTACCTGCACTGGCGCGCTCTGGCGCCCTCTGCACCGATGGCCTCTGTCTAAGTCCGGACCA	300
OY	397	CGTCAATTGCTTACGAGACCGCTCCAGACCGCGCGAGCGGCGCACTACCCCTGCTTGGCCCCGA	456
Db	301	CGTCAATCGCTTACGAGACCGGTGCAGACCGCGCGAGCGGCGCACTTCCGCTCTGGCCCCGA	360
OY	457	TGAGCGAGTGGCGGGTGGAGTTCGCCCCCGCCAGGTGGCGGTGTACCACTGTATGGAGACC	516
Db	361	TGAGCGAGTGGCGGGTGGAGTTCGCCCCCGCCAGGTGGCGGTGTACCACTGTATGGAGACC	420
OY	517	AAGGGGCGCGGGTGTGCTGATGGGCGGGGGTGCCTGGCGGTGCAACCGCGGACGCTGTGG	576
Db	421	ACGGGCGGTGGCGGGCTGTGTGATGGCGGGCGTCCCGCGGTCAACCTTCGCGACGCTGTGG	480
OY	577	TGATGGGCGCGCGCACCCCGCGGTACACAGCGAGCCGCGATCGCCCAAGGCGATGGGCGCGA	636
Db	481	TGATGGGCGCGCGCGCACCCCGCGGTACACAGCGCGCGCGGTGCCCAAGGCGATGGGCGCGA	540
OY	637	CCGTTACGGTCTTACACATCAACATCGACAACATTTGGCAACTGAGAGCGGAGTTCTGGC	696
Db	541	TGGTACCGGTGTGATGTCAACATCAACAGCTTCGCGCAATGACAGCGGAGTTCTGGCG	600
OY	697	GCCGGATCCACACTGCTACTATCATGAGCTTACGAGAGTGAAGGGTGCAGGTCAACAGTCCGG	756
Db	601	GTCGGGTCCGAGACCGGTACTGTGTGACCTCTGACACTGAGAGATGGCGGAGTCCAGCGCG	660
OY	757	ACCTGGTATTTGGGCGCGTCTT	778
Db	661	ACATGGTATGGGGCGCGTCTT	882

Matches 604; Conservative 0; Mismatches 457; Indels 13; Gaps 5;

```

QY 61 ATGCGCGTGGTATTCGACCGAGACCAAAACCAATTCGGGGTGGCCATCACC 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 ATGAGATCGGATTCGCAAAAGAAATCAAAACATGACCGGCTGCACACCG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 GCCGGCGTCCGGAATACCCGTCGTGGCCATGAGTGGT---CATCCAGCGAGTGGC 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GCAGGCGTATGAGCGCTCGCAAAAGCGGGCATGAGTGGTATGATGAGAGCGAAGGC 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 178 GGAAGGCGTGGCTATCACCAGCGGATTCAGAGCGGCGGCGCACTGGTGGC 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GCGGCTGGTGGGGTTCCTCCGATTCGAGTAAAGAAACCGGCGAGCTGATCGTTGC 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 238 ACCGCCACAGGTGTGGGCGGAGCGCTGATTCCTCAAGTCAGAAAGACGATGGC 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 CGAAGCTGAGAGATGCTTGGACGGCGAGATGTTGAAAGGAAAGAGCGCGTGGCT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 298 GCGGAATATCAGCGCGTGCACACAGG-GCGATCTTGTACGTTCTTCATTGGCGCG 356
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 CGAGAGTCCGCTATTTCCCGCGGATGATTTGTTGATGATTTGATTTGATTTGATTTG 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 357 TCAGTGTCTTCACCGATGCGTGTGGATTCGCGACACGATTCATTTGCTACGAGACC 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GCGGAACGCGTACAGAAAGCGTGTGACAGCAAAAGTGGTGGCGATGCTTACGAGACG 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 417 GTCCAGACCCCGACGGCGGCACTACCTGCTTCCCGCATGAGCGAAGTCG GGTGCA 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 GTGAGTGTGGGAGAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 477 CTGCGCCCGGAGTGTGGCTTACCACTGATGAGCAACCAAGGGGCGCGGTGGTGGT 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 ATGTGCGTGAATGCGCGCGCCCACTTCTCGAGAAAGCGGAGGAGGAGGAGGATTTTG 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 537 ATGGCGGGTGGCGCGGCTGCAACCGCGGCGCTGTGTATCGGCGCGCGCACCGCC 596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 CTGGCGCGGCTGGCGGAGTGGCGCGCGGCGGCAAAATGACATGACGCGCGGAGCGCG 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 597 GCGTACAGCCAGCCCGCATCGCAAGGAGTGGGCGGAGCGGTACGGTTTACAGATC 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 GGAAGCAAGCGGAGAAATCGGCTGCTGCGGTGACAGCGGTGAGATTTGGACATTT 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 657 AACATGACAACTCGGCAACTCGAGCGGAGTTCGCGCGCGGATTCACACTCGCTAC 716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 601 AACGCGGAGCGGCTGCGCGGAGCTGATGATTTGTTCCGCAACCGTACAGAGCGCTATG 660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 717 TCATCGGCTACGAGCTGAGAGGTGCGGTCAAAACGTCGACCTGATGGTGGGCGGTC 776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 TCCAACTCGTACCATATGCGCGAGTGGTGGCGGATTCGATTTGGTGGTGGCGGTC 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 777 CTGGTGCAGAGCGCGCAAGGACCAATTTAGTCTCGAATTCATCTTCGCGCATATGAA 836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 TTGATCCCGGGGCGGAA--AGCGAAGCTGTGACGGAAGAGATGCGCTCGATACG 777
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 837 CCAAGTGGTACGAGGATATAGCATGAGACGAGGCGGCTTTTGA--GAGTCA 893
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 778 CCGGAGATCGGTGTGTGCTGATATGCGCATTTGACCAAGCGGATTTTGAAGAGCAGC 837
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 894 CGACCGACCACTAGACCAACCGGAGCTTGCCTGACAGACAGCTGTTTACTGGTG 953
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 838 CGACCGTACGACGAGATCGCATACGATCAAGCAGCGCGTCCATTTAGCGCGTC 897
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 954 GCGAATACGCGCGCTGCGGAGCAAGAGCTGACGCTACGCGTACCAAGCGAGCATG 1013
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 898 GCGCAACTTGGCGG---GTGCGCGGACGTCGATTCGCGCTTACGAAGCTACGATC 954
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1014 CCGTATGTGTCGAGCTTGGCGGACATGCTGGCGGCGCGCTCGCATTCGCGCA 1073
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 955 CCAATACCGCTTGCAAAATCGCCAAACAAAGGCTACCGCGCGGTTCTTGATTAACCGGCG 1014
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1074 CTAGCCAAAGGTCTTTGAGAGCAGAAAGGCGGTACTGTTCGAAAGGCTGGC 1127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1015 CTGTTAAAGGATCAACAGGCTGACGGGACATCGTGTACGAAGGCTCGC 1068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6  
V52155/c  
ID V52155 standard; DNA; 28171 BP.  
AC V52155;  
DT 23-OCT-1998 (first entry)  
DE Streptococcus pneumoniae genome fragment SEQ ID NO:22.  
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
computer readable medium; vaccine; pharmaceutical composition; ds.  
PN Streptococcus pneumoniae.  
PM W0918931-A2.  
PD 07-MAY-1998.  
PF 30-OCT-1997; U19588.  
PI 31-OCT-1996; U502960.  
PI (HUMAN) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,  
PI Kunsch CA, Rosen CA;  
DR MPI: 98-27225/24.  
PT Computer-readable medium with recorded Streptococcus pneumoniae  
polynucleotide sequences - useful in diagnostic kits and assays, and  
PT pharmaceutical compositions and vaccines for Streptococcus  
PT pneumoniae.  
PS Claim 1; Page 273-289; 14099p; English.  
CC The present invention describes a computer readable medium which has  
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
CC on it, or a representative fragment or a sequence at least 95% identical  
CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
CC pneumoniae. The present invention also describes an isolated nucleic acid  
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae  
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
CC by a process comprising: (a) screening a genomic DNA library using as a  
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
CC to 391, identifying members of the library which contain sequences  
CC that hybridize to the target sequence and isolating the nucleic acid  
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
CC from an organism, amplifying nucleic acid molecules whose nucleotide  
CC sequence is homologous to amplification primers derived from the  
CC fragment of the S. pneumoniae genome to prime the amplification and  
CC isolating the amplified sequences. The computer readable medium can be  
CC used in a computer-based system for identifying fragments of the  
CC S. pneumoniae genome of commercial importance, or expression modulating  
CC fragments of the S. pneumoniae genome. Products from the present  
CC invention can be used in diagnostic kits and assays, and pharmaceutical  
CC compositions and vaccines for S. pneumoniae.  
SQ Sequence 28171 BP; 9507 A; 5080 C; 6440 G; 8144 T;

Query Match 16.4%; Score 202; DB 1; Length 28171;  
Best Local Similarity 53.9%; Pred. No. 16-33;  
Matches 533; Conservative 0; Mismatches 435; Indels 20; Gaps 5;

```

QY 61 ATGCGCGTGGTATTCGACCGAGACCAAAACCAAGATTCGGGTGGCCATCACC 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21711 ATGTTATTCGAAATCCCAABAGAAATTAATAAGAAACCGTTCGCGCTCACACT 21652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 GCCGGCGTCCGGAATACCCGTCGTGGCCATGAGTGGTCTCCAGCAGAGTGGCGGA 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21651 GCAGGTTTCATAGCTTAGTACTGTGTCATCGTGTCTTATGCAAAACAAATGCTGT 21592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 181 GAGGCGTGGCTTACCAACCGGAGTTTCAGGCGGACCGCGCACTGGTGGAGAG 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21591 CTGGGTCTGCTTACTGATGCTGACTATCAAAAGCAAGAGCTGAGTTGGCTACT 21532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 241 GCCGACAGGTGTGGCGCGAGCTGATTTATGCTCAAGGTCAAAGACGATAGCGGCG 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21531 GCTGTGTAACCTTGCGC---AGCAGAGTTGTGTGAAGTAAGATCTTAAGTTCT 21475
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 301 GAATAGCGCGCGCTGCGACACGGGC-GATCTGTTCACGTTCTTGATTTGGCGCGTCA 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21474 GAATAGGTTACTTGGCGGACGATCTTCTTCTTACCACTTGCACATGCGCGCTGT 21415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

OY 360 CGTGGTCCAGCAGATCGTGTGATTCGGCAGCAGCAGTCAATTCGACAGACCGTC 419
DB 21414 CCAGATTTAGCAGATCTATGTT-----AACAGCAAAAACACTGAACTGTT 21367
OY 420 CAGACCCGCCGAGCGCAGTACCCCTGCTTGGCCCGATAGCGAACTGCGCGGAGCTC 479
DB 21366 CCGTGAACATCAAGAGCAACATCCGCTCTGCTTCCATGATGAGTGTGACAGTGTATG 21307
OY 480 GCGCGCCAGTGGCGGCTGACCACTGATGCGCAACCCAGGGGCGCGCGCTGCTGATG 539
DB 21306 GCGTTCATTAATGAGAGCTACTCTTACTTAAGCAAGCTGTGGCTGCTGCTTACTT 21247
OY 540 GCGCGGCTGCGCGCTGCAACCGCGCAGCTGCTGATGGCGCGCGCAGCGCGC 599
DB 21246 GCGTGGTACAGAGTGTTCACAAAGGAAAGTAACATATGATGCTGCTGCTGCT 21187
OY 600 TACACGACCGCCGATGCGCAACGCGATGGCGCGCAGCGCTTACAGCTTACAC 659
DB 21186 ACACATGCTGCGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 21127
OY 660 ATCGACAACTTGGCAGCTGCGCAGCGCTGCTGCGCGCGCTGCTGCTGCTGCTGCT 719
DB 21126 TCGAACGCTTCTCACTTACAGAGCTTTCGAGATTAATTCACCTTATGCT 21067
OY 720 TCGGCTTACAGCTGAGGCTGCGCTCAACGCTGCGCAGCTGCTGCTGCTGCTGCTG 779
DB 21066 AATTCATTAACATTAAGCAAGTGTGAGAGATGCTGATGCTGATGCTGCTGCTGCT 21007
OY 780 GCGCAGCGCGCAAGCAGCAATTAATGCTGATTCCTGCTGCTGCTGCTGCTGCTGCT 839
DB 21006 ATCCCTGCGGCAAAAGCAGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20947
OY 840 GGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 396
DB 20946 GCGCTCTGTA-TGCTTACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20888
OY 897 CCGACACCTTACAGCAGCAGCGAGCTTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 956
DB 20887 GTCACACGCAAGTATGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20828
OY 957 AACATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1016
DB 20827 AATATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20768
OY 1017 TATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1044
DB 20767 TATATGAAAGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20740

RESULT 7
V74442/c
ID V74442 standard; DNA: 9280 BP.
AC V74442:
DE 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #131.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc-feature 841..900
FT /tag- a
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature 2641..2700
FT /tag- b
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"

```

```

FT misc-feature 4441..4500
FT /tag- c
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature 6241..6300
FT /tag- d
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature 8041..8100
FT /tag- e
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT EP-786519-A2.
FT 30-JUL-1997.
FT 07-JAN-1997; 100117.
FT 05-JAN-1996; US-009861.
FT (HOMA-) HUMAN GENOME SCI INC.
FT Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
FT Rosen CA;
FT WPI: 97-374922/35.
FT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
FT stored on computer readable medium and used in the production of
FT anti-S.aureus vaccines
FT Claim 1: Page 710-715; 3271pp; English.
FT This sequence represents one of 5191 Staphylococcus aureus DNA sequences
FT of the invention. The DNA sequences are recorded on a computer readable
FT medium, preferably selected from a floppy or hard disk, random access
FT memory (RAM), read only memory (ROM) or CD-ROM. Homology searches using
FT the S.aureus DNA sequences allows putative functions to be assigned so
FT that protein-encoding or regulatory regions of commercial, therapeutic or
FT industrial importance can be obtained. Specifically, sequences which are
FT likely to encode antigens have been identified and these polypeptides can
FT be used in a vaccine composition against S.aureus infection. The
FT polypeptides can also be used in a kit for the immunodetection of
FT S.aureus in a sample. S.aureus is implicated in numerous human diseases,
FT including cellulitis, eyelid infections, food poisoning, osteomyelitis,
FT skin and surgical wound infections, scalded skin syndrome, toxic shock
FT syndrome, etc. Organisms transformed with the DNA sequences can be used
FT for recombinant production of the polypeptides. The new DNA sequences
FT (and their fragments) are useful as primers or probes for isolating
FT homologues of any of the S.aureus DNA sequences contained on the
FT computer readable medium.
FT Sequence 9280 BP; 2958 A; 1649 C; 1309 G; 3060 T;

Query Match 13.6%; Score 168; DB 1; Length 9280;
Best Local Similarity 50.6%; Pred. No. 1e-26;
Matches 484; Conservative 0; Mismatches 465; Indels 7; Gaps 3;

OY 151 CATGAGTGCATACACGAGCGGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 210
DB 9280 CATACTGTTTATGTCGAAACAATGCGGTTGACGATCTTCTTGAAGATGATGATAC 9221
OY 211 AAGCGGAGCGCGCGCACTGCTGCGACCGCGCAGCAGTGTGGCGGAGCGCTGATTA 270
DB 9220 AAGAGAGCAGGTGCTGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9164
OY 271 TTGCTCAAGTCAAGAACGATAGCGGGAATAGCGCGCTGCGACAGCGGC-GATC 329
DB 9163 GTTATTAAGTTAAAGCACTGGAATCTGAATATCAATTTTAAAGAGGCTGTA 1014
OY 330 TTGTCACGCTCTTGCATTTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 389
DB 9103 TTATTCATCTTATCTTCAATTTAGCAATGAAGAAAATTAACACAGCTTTGATGATGA 9044
OY 390 GGCACACGTAATGCTTACAGACGCTCAGACCGCGCGGCGGAGCTACCTCTCTT 449

```

Db 9043 AAGTAATTAGTATGATGAGACTGTGCACTTACCAACCGACTTTTACCATTTGTA 8984  
 QY 450 GCGCCGATGAGCAAGTCCGCGGCTGCTGCGCCGAGGTTGGCTTACACCTGATG 509  
 Db 8983 TCACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8924  
 QY 510 CGAACCCAGAGGAGGCGGCTGTGTGATGATGATGATGATGATGATGATGATGAT 569  
 Db 8923 AACTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8864  
 QY 570 GTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 529  
 Db 8863 GTAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8804  
 QY 630 GCGCGAGCTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689  
 Db 8803 GGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8744  
 QY 690 TTCTGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 749  
 Db 8743 TTCTGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8684  
 QY 750 GGTGCGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 809  
 Db 8683 CAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8624  
 QY 810 TCGAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 869  
 Db 8623 ACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8564  
 QY 870 CAGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 926  
 Db 8563 CAGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8504  
 QY 927 GTGACGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 986  
 Db 8503 AAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8444  
 QY 987 ACCTACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1046  
 Db 8443 ACCTACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8384  
 QY 1047 CCGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1102  
 Db 8383 AGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8328

RESULT 8  
 ID T67971 standard; DNA: 1074 BP.  
 AC T67971.

DT 15-JUL-1997 (first entry)  
 DE H. pylori membrane protein ORF 05cp20518orf61.  
 KW Vaccines; prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;  
 KW membrane; amino acid; metabolism; ds.  
 OS Helicobacter pylori.  
 FH Key Location/Qualifiers  
 FT 1..1074  
 FT /tag-  
 FT /note- "no stop codon given"  
 PD 19-DEC-1996  
 PF 06-JUN-1996: U09122.  
 PR 07-JUN-1995: US-487032.  
 PR 01-APR-1996: US-630405.  
 PA (Aspr) AsprA AB.  
 PI Berglindh OT, Smith D, Mellgaard BL.  
 DR WPI: 97-052306/05.  
 DR P-PDB: W20718.  
 PT Helicobacter pylori nucleic acid sequences and related  
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori

PT Infection, and to detect Helicobacter  
 PS Claim 27; Page 819; 1481pp; English.  
 CC The present sequence encodes a Helicobacter pylori membrane  
 CC protein likely to contain four membrane spanning regions.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors.  
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the  
 CC bacterial DNA. The sequences were analysed for ORF of at least 180  
 CC nucleotides, and the predicted coding regions defined by computer  
 CC evaluation. To identify likely H. pylori antigens for vaccine  
 CC development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences  
 CC of interest, particular regions can be isolated from H. pylori by  
 CC PCR amplification for recombinant polypeptide production, e.g. in  
 CC E. coli hosts.  
 SQ Sequence 1074 BP; 302 A; 189 C; 297 G; 286 T;

Query Match 7.1%; Score 87.6; DB 1; Length 1074;  
 Best Local Similarity 46.4%; Pred. No. 4.2e-10;  
 Matches 372; Conservative 0; Mismatches 414; Indels 16; Gaps 2;

QY 265 GATTATTGCTCAAGTCAAAAGACGATAGCGCGGATAGCGCGCTGCGA-CACGG 323  
 Db 133 GATTGTTGCTCAAAAGACGATAGCGCGGATAGCGCGCTGCGA-CACGG 192  
 QY 324 GCGATGTTGCTCAAGTCAAAAGACGATAGCGCGGATAGCGCGCTGCGA-CACGG 383  
 Db 193 GCGACTGTTGTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 252  
 QY 384 GATTCCGCGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 443  
 Db 253 AATTAATAAATCACTTCTATTGCTACGTAACCATGTCGCGGCTTAATAACACTACCT 312  
 QY 444 CTGCTGCGCGGATGAGAC AGTCGCGGCTGACGTCGCGCGGCTGCGA-CACGG 495  
 Db 313 ATTTAGCGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 372  
 QY 495 -----GCTTACCACTGATGCGAACCAAGGCGCGGCTGCTGATGCGGCGGCTG 548  
 Db 373 TTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 432  
 QY 549 CCGGCGGCTGACCG 608  
 Db 433 TCGGCTGCGCAAGGCGCTAATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 492  
 QY 609 GCGCGATGCGCAAGGCG 668  
 Db 493 GCGCAAGCTTTAAGCAATGCG 552  
 QY 669 CTTCGCACTGACG 728  
 Db 553 TTACAAACACACCTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 612  
 QY 729 GAGCTGAGGCTGCGCAAGGCG 788  
 Db 613 AATATCATTCAGCTTAAAGCG 672  
 QY 789 GCGAAGCGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 848  
 Db 673 AGCCAAACCCCTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 732  
 QY 849 CTGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 908  
 Db 733 GTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 792  
 QY 909 GACCAAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 968  
 Db 793 TCTAACCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 852



OY 969 TCGTGCCGGAAGCTGACCTAGCGCTGACCAACGACGATGCCGTATGCTCGAG 1028  
 DB 853 ATGTCTGCTTAAGACGACCTAGCGCTTATAGCATGCGAGTGGCGGTATTTGTTAT 912  
 OY 1029 CTGCGCAGCATGGCTGGCGG 1050  
 DB 913 TATTAGCATGCTTGAAG 934  
 RESULT 9  
 V43039/c  
 ID V43039 standard; DNA; 544 BP.  
 AC V43039;  
 DT 09-NOV-1998 (first entry)  
 DE Streptococcus pneumoniae polypeptide coding region.  
 KW Polypeptide; ORF; open reading frame; infection; bacterial;  
 streptococcal; bacteremia; diagnosis; prophylaxis; ds.  
 OS Streptococcus pneumoniae.  
 FH Location/Qualifiers  
 FT Key complement (39..334)  
 FT CDS  
 FT /tag- a  
 FT /note- "polypeptide"  
 PN W09823631-A1.  
 PD 04-JUN-1998  
 PF 24-NOV-1997; U21976  
 PR 27-NOV-1996; US-031879.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO,  
 PI Reid RH, Zarfos PN;  
 PI WPI: 98-322654/28.  
 DR P-PSDB: W62769.  
 DR Streptococcus pneumoniae polynucleotides - useful for developing  
 PT products for diagnosis, prevention and treatment of infections e.g.  
 PT pneumonia, bacteremia, meningitis or endocarditis  
 PS Claim 1: Page 165-166; 181pp: English.  
 CC The sequence is that of a Streptococcal polypeptide coding region.  
 CC The polypeptide can potentially be used for the diagnosis and  
 CC prevention of bacterial infections, especially SP infection.  
 CC It may be used for the treatment of diseases such as otitis media,  
 CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural  
 CC empyema, endocarditis or infection of the cerebrospinal fluid.  
 CC Sequence 544 BP; 149 A; 121 C; 124 G; 150 T;  
 SQ  
 Query Match 5.68; Score 69.2; DB 1; Length 544;  
 Best Local Similarity 53.48; Pred. No. 2.6e-06;  
 Matches 191; Conservative 0; Mismatches 163; Indels 4; Gaps 2;  
 OY 690 TTCTGCGCGCGGATCCACACTGCTACTCATCGGCTACGAGCTCGAGGTGCCCTCAA 749  
 DB 543 TTGGAGCAAAATTCATTTCTTATCTATTCATTCACATGAGCAAGTATGAGA 484  
 OY 750 CGTGGCGAGCTGTATTTGGGGCGTCTGCTGCGCAGCGCCCAAGCAATTAATC 809  
 DB 483 GATGCGATGTGTGATTTGGACCATTCATCCCTGGGCAAAAGCACCGGAATGGTG 424  
 OY 810 TGGAAATTCCTGTGGCGCATATGAACAGGTGGCTACTGGTATATAGCATGAC 869  
 DB 423 ACGATGAGATGTGTAACAAATGCGTCCAGGCTCTGA-TGTTGACGTTGCTTTGAC 365  
 OY 870 CAGGGCGGCTGTTTGA--AGGCTCAGACCGACCACTACTGACACCCGCACTTCGCC 926  
 DB 364 CAAAGTGGCGTTATGCAAAACAGCTGACGCTGTGACACGACGATGAAACCCGCTATGAA 305  
 OY 927 GTGACGACAGCTGTATTACGCGGTGGGAGACATGCCGCTCGTGGCGGAAGAGCTG 986  
 DB 304 AAACAGGTTGTTCTCCATATGCCGTTGGCAATATCCCTGGTGGCGGTTGCCGCACTTCA 245  
 OY 987 ACCTACGCGCTGACCAACGACGATGCGTATGTGCTGCGTGTGCGGACCATGGCT 1044  
 DB 244 ACCATGCGCCCTAACCAATGTACTCTTCTTATATGGAAGCTTTGGCTGCAAGGAT 187

RESULT 10  
 ID V74370  
 V74370 standard; DNA; 31096 BP.  
 AC V74370;  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SEQ ID #59.  
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.  
 OS Staphylococcus aureus.  
 FH Location/Qualifiers  
 FT Key  
 FT misc-feature  
 FT 1201..110  
 FT /tag- a  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature  
 FT 3001..3060  
 FT /tag- b  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature  
 FT 4801..4860  
 FT /tag- c  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature  
 FT 6601..6660  
 FT /tag- d  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature  
 FT 8401..8460  
 FT /tag- e  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature  
 FT 10201..10260  
 FT /tag- f  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature  
 FT 12001..12050  
 FT /tag- g  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature  
 FT 13801..13860  
 FT /tag- h  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature  
 FT 15601..15660  
 FT /tag- i  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature  
 FT 17401..17460  
 FT /tag- j  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"





[illegible]

RESULT 12  
 036859/c  
 ID 036859; standard; DNA; 390 BP.  
 AC 036859;  
 DT 22-JUN-1993 (first entry)  
 DE PCR primer for 5' fixed sequence contg. T7 promoter and RBS.  
 KW Systematic peptide evolution by reverse translation; SPERT; ligand;  
 OS specific; inhibitors; probes; assay; cell sorting; ss.  
 PN Synthetic.  
 PN W030303172-A.  
 PD 18-FEB-1993.  
 PF 31-JAN-1992; 000801.  
 PR 01-AUG-1991; US-739055.  
 PR (UYRE-) UNIV RES CORP.  
 PI Gold L, Pribnow D, Smith JD, Tuerk C;  
 WP1: 93-076529/09.  
 FT Systematic polypeptide evolution by reverse translation - used  
 PT for prodn. of polypeptide ligand specific for desired target  
 PT molecule  
 PS Example 1: Page 84; 98pp; English.  
 CC SPERT is used to select novel polypeptides that bind the antibody  
 CC of the epitope commonly recognised by the antisera from autoimmune  
 CC mice which are the F1 progeny of a cross of NZB and NZW parents  
 CC (Portanova et al., J. Immunol. 144, 4633, 1990). The known epitope  
 CC consists of ca. 10 amino acids at the N-terminus of the histone H2B  
 CC protein. To make mRNA encoding candidate polypeptides a 5' fixed  
 CC sequence composed of a T7 promoter sequence and a ribosome binding  
 CC site which is recognised by both prokaryotic and eukaryotic ribosomes,  
 CC terminating in a restriction endonuclease site is synthesised and cloned  
 CC using a number of oligonucleotides (example shown). A 3' fixed sequence  
 CC is placed into a restriction site to provide an mRNA encoding the C-  
 CC terminal trailer sequence of ca. 100 nucleotides lacking stop codons.  
 CC In addition, a 3' primer annealing site is provided so that cDNA  
 CC synthesis can be accomplished on the mRNA recovered from partitioned  
 CC ribosome complexes. See also Q36845-63.  
 SQ Sequence 390 BP; 125 A; 126 G; 133 G; 6 T;

Query Match	4.1%	Score 50.8;	DB 1;	Length 390;
Best Local Similarity	46.28;	Pred. No. 0.017;		
Matches 169;	Conservative	0;	Mismatches 197;	Indels 0;
			Gaps	0;

[illegible][illegible]

ID	Accession	Standard	DNA	Length (BP)
RESULT 13	T68715			
AC	T68715	standard	DNA	15872 BP.
DT	01-SEP-1997	(first entry)		
DE	Streptomyces venezuelae polyketide synthase vep ORF1.			
KW	Polyketide synthase; polyhydroxyalkanoate monomer synthase;			
KW	polyhydroxybutyrate; biodegradable polymer; vep gene;			
KW	metabolic engineering; ss.			
OS	Streptomyces venezuelae.			
PH	key	Location/Qualifiers		
FT	cds	20..13912		
FT	cds	/*tag- a		
FT	cds	14056..14136		
FT	cds	/*tag- b		
FT	cds	14148..15827		
FT	cds	/*tag- c		
PN	M09722711-A1.			
PD	26-JUN-1997.			
PF	18-DEC-1996:	020119.		
PR	19-DEC-1995:	DS-008847.		
PA	(MINU )	UNIV MINNESOTA.		
PI	Sherman DH, Williams MD, Xue Y:			
DR	WPI: 97-241701/31.			
DR	P-PSDB: M19629-30.	AND M00918.		
PT	Expression cassettes for production of polyhydroxyalkanoate(s) -			
PT	provide wide range of biodegradable polymers for medical or			
PT	industrial use			
PS	Claim 54; Fig 23; 91pp; English.			
CC	Streptomyces venezuelae vep ORF1 (T68715) comprises the polyketide			
CC	synthase (PKS) gene cluster encoding a polypeptide of 12 carbons (see			
CC	also M19629-30 and M00918). It contains 5 PKS modules, with a 5'			
CC	loading module and a 3' end domain. Each of the sequenced modules			
CC	includes a keto-ADP, an acyltransferase, a dehydratase, a keto-			
CC	reductase and an acyl carrier protein domain. The gene cluster was			
CC	cloned using a heterologous hybridisation strategy from a genomic			
CC	DNA library. A novel expression cassette encoding the first module			
CC	from the vep gene cluster and module 7 from the Streptomyces tylo			
CC	gene cluster has polyhydroxyalkanoate (PHA) monomer synthase			
CC	activity and can be used for PHA prodn. in host (esp. insect) cells			
CC	for use as a biodegradable polymer.			
CC	Sequence 15872 BP: 2093 A: 5503 G: 1963 T:			
CC	50			

Query Match	4.1%	Score 50.4	DB 1	Length 15872
Best Local Similarity	47.7%	Pred. No. 0.026		
Matches 187	Conservative 0	Mismatches 196	Indels 9	Gaps 1
Qy	415	CGTCCAGCCGGCCGACGGCGCACTACCCCTCTTCCCGCATGAGCAAGTCGGCGATCG	475	
Db	1540	CGTCTCGAGCGGCTCTGCGCCCTCGCCGACGGAGCGCCCGCCGCGGAGTGTATCCGG	1599	

QY 476 ACTGCGCCGCAAGGTGGGGCTTACCAACGATGGAACCAAGGGGGCGGGTGTCT 535  
 DB 1600 CACCCCGCCCGCCGCGCTGCGCTCTCTACAGGCGAGGCCACATGACGG 1659  
 QY 536 GATGGCGGGGTGCGCGGCGTCAACCGGCGAGCTGTGTGTATGGCGCGCACCG 595  
 DB 1660 CATGGCATGAGTTTATAGCGCCGACCGCGCTTCGGAGGCGCTTGAGCGCTGCG 1719  
 QY 596 CGGCTACAAACGACCGCGCATGCGCAAGGCGATGGCGCGACCGTTAGGTT----- 648  
 DB 1720 CGCGCACTGAGACCGCTCTGACCGCGCGCGCGCAACTGTCTGCGCGCGCGGCGACAC 1779  
 QY 648 -CTAGCATCAACATGCAAACTTGGGACATCGACGGGAGTTGTGGGCGCGATCCA 706  
 DB 1780 CCTCGACCGGACGCTCAACACACCGCGCGCTTGTGGCGGTGAGAGTGGCGCTCCACCG 1839  
 QY 707 CACTCGCTACTATCGGCTACGAGCTCGAGGGTGGCTCAACGTCGACCTGTGTAT 766  
 DB 1840 CCTCGTGAAGTCTGGGGGGTACGCCGACGTCGTCGCGGCGCACTCGGTGGCGAGAT 1899  
 QY 767 TGGGGCGCTGCTGGTGGCGACGCGCCACAGCGAC 798  
 DB 1900 CAGCGCGCCGACGTCGCGGGGTCTGTCTGC 1931

## RESULT 14

V4433/c  
 ID V4433 standard; DNA: 985 BP.  
 AC V4433;  
 DT 09-NOV-1998 (first entry)  
 DE Mycobacterium tuberculosis antigen Xp25 3' DNA.  
 KM Tuberculosis; infection; diagnosis; antigen; Xp25; ss.  
 OS Mycobacterium tuberculosis strain Erdman.  
 PN W09816645-A2.  
 PD 23-APR-1998.  
 PF 07-OCT-1997: U18214.  
 PR 13-MAR-1997: US-818111.  
 PR 11-OCT-1996: US-72967.  
 PA (CORI-) CORIXA CORP.  
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedyck TS; WPI: 98-251292/22.  
 DR WPI: 98-251292/22.  
 PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used to develop products for the detection of M. tuberculosis infection  
 PT and diagnosis of tuberculosis  
 PS Claim 4; Page 187-188; 250pp; English.  
 CC This is the 3' region of DNA coding for an antigenic portion of Mycobacterium tuberculosis antigen Xp25; 5' DNA is provided in V44438. Xp25 DNA was isolated from a M. tuberculosis strain Erdman genomic DNA expression library using sera from patients having extrapulmonary tuberculosis. It bears no similarity to known sequences. The invention relates to methods and compositions for diagnosing tuberculosis. It provides polypeptides (see W64491-W6479) comprising an antigenic portion of a soluble M. tuberculosis antigen, or an immunogenic portion of a M. tuberculosis antigen, as well as DNA sequences encoding such polypeptides, recombinant expression vectors and transformed or transfected host cells. Also claimed are methods and diagnostic kits for detecting M. tuberculosis infection in a patient using these polypeptides, antibodies or oligonucleotide probes and primers.  
 CC primers.  
 SQ Sequence 985 BP; 138 A; 330 C; 423 G; 94 T;

Query Match 3.8%; Score 46.4; DB 1; Length 985;  
 Best Local Similarity 48.8%; Pred. No. 0.14;  
 Matches 125; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 384 GATTCGGGACACGATTCATTCCTACGAGAGCGTCCAGACGCGGACGCGCATACCC 443  
 DB 351 GAGGCCGAGGCGGACGCGCTGCGCCCGCATGCGCGCGGCGCGCGCGCGCGCC 292

QY 444 CTGCTGCCCGATGAGCGAGTCGCCGCTGACCTCGGCCCGCCAGGTTGGCGCTACAC 503  
 DB 291 GCCGATGCTCTGACACTACCGTGGCGCGGACAAAGCGCGGTGCGCGGTACCGGTGC 232  
 QY 504 CTGATGCGAACCAGGGGGCGCGGTGTGTGATGAGGGGGGTGCGCGGTGCGAACC 563  
 DB 231 CCG 172  
 QY 564 GCCGAGTGTGTGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623  
 DB 171 AGCACCAGGTGGGTGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 112  
 QY 624 GGCATGCG 639  
 DB 111 TGCACCG 96

## RESULT 15

V64548/c  
 ID V64548 standard; DNA: 985 BP.  
 AC V64548;  
 DT 27-JAN-1999 (first entry)  
 DE M. tuberculosis immunogenic polypeptide Xp25 3'-end DNA.  
 KM Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB; vaccine; pharmaceutical; infection; diagnosis; ss.  
 OS Mycobacterium tuberculosis.  
 PN W09816646-A2.  
 PD 23-APR-1998.  
 PF 07-OCT-1997: U18293.  
 PR 13-MAR-1997: US-818112.  
 PR 11-OCT-1996: US-730510.  
 PA (CORI-) CORIXA CORP.  
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedyck TS; WPI: 98-261042/23.  
 DR WPI: 98-261042/23.  
 PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to develop products for the detection of M. tuberculosis infection  
 PT and for diagnosis, treatment and prevention of tuberculosis  
 PS Claim 4; Page 176; 230pp; English.  
 CC This sequence encodes an immunogenic portion of a soluble Mycobacterium tuberculosis (MT) antigen which can be used in a method for inducing protective immunity against tuberculosis (TB). This sequence can be formulated into vaccines and/or pharmaceutical compositions for immunizing against M. tuberculosis infection or may be used for the diagnosis of tuberculosis.  
 CC immunizing against M. tuberculosis infection or may be used for the diagnosis of tuberculosis.  
 SQ Sequence 985 BP; 138 A; 330 C; 423 G; 94 T;

Query Match 3.8%; Score 46.4; DB 1; Length 985;  
 Best Local Similarity 48.8%; Pred. No. 0.14;  
 Matches 125; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 384 GATTCGGGACACGATTCATTCCTACGAGAGCGTCCAGACGCGGACGCGCATACCC 443  
 DB 351 GAGGCCGAGGCGGACGCGCTGCGCCCGCATGCGCGCGGCGCGCGCGCGCGCGCC 292  
 QY 444 CTGCTGCCCGATGAGCGAGTCCGCGTGCACGCGCGCGCGCGCGCGCGCGCGCGCG 503  
 DB 291 GCCGATGCTCTGACACTACCGTGGCGCGGACAAAGCGCGGTGCGCGGTGCGAACC 232  
 QY 504 CTGATGCGAACCAGGGGGCGCGGTGTGTGATGAGGGGGGTGCGCGGTGCGAACC 563  
 DB 231 CCG 172  
 QY 564 GCCGAGTGTGTGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623  
 DB 171 AGCACCAGGTGGGTGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 112  
 QY 624 GGCATGCG 639  
 DB 111 TGCACCG 96

Fri Jun 23 09:31:34 2000

us-09-362-485-6.rng

Page 12

Search completed: June 22, 2000, 15:09:11  
Job time: 17877 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:08:43 ; Search time 341.15 Seconds  
(without alignments)  
470,560 Million cell updates/sec

Title: US-09-362-485-6  
Perfect score: 1235  
Sequence: 1 ATCTTGACAGTTATATGAC.....GGGAGAGCATGATGCGCC 1235

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 6492525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/PCRTS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/1na/Dackfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	3.7	2582	4	US-08-816-105A-2
2	45.4	3.7	833	3	US-08-403-852D-3
3	45.4	3.7	5392	3	US-08-403-852D-1
4	45.4	3.5	78958	1	US-08-258-261B-6
5	43.6	3.5	28958	1	US-08-456-837-6
6	43.6	3.5	28958	1	US-08-457-646A-6
7	43.6	3.5	28958	1	US-08-458-076A-6
8	43.6	3.5	28958	2	US-08-764-233A-6
9	43.6	3.5	28958	2	US-08-457-335A-6
10	43.6	3.5	28958	2	US-08-429-214-6
11	43.6	3.5	49377	2	US-08-764-233A-1
12	43.6	3.4	1833	3	US-08-403-852D-6
13	42.4	3.4	2185	1	US-08-173-508-3
14	42.4	3.4	2185	3	US-08-265-310-3
15	41.2	3.3	2414	7	US-08-461-775-10
16	41.2	3.3	1650	3	US-07-642-734C-3
17	41.2	3.3	20235	2	US-08-804-227C-7
18	40.8	3.3	44377	4	US-08-804-198-1
19	40.8	3.3	44377	4	US-08-804-473-1
20	40.6	3.3	12001	2	US-08-458-568A-11
21	40.6	3.3	12001	2	US-08-343-428-1
22	40.2	3.3	11219	2	US-07-642-734C-1
23	40.2	3.2	474	3	US-08-403-852D-14
24	39.6	3.2	459	4	US-08-387-942C-35
25	39.4	3.2	1998	7	5212296-8
26	39.4	3.2	1998	7	5212296-8
27	39.4	3.2	1998	7	5212296-8

28	39.4	3.2	2668	3	US-08-461-775-11	Sequence 11, Appl
29	39.4	3.2	12588	4	US-08-387-942C-1	Sequence 1, Appl
30	39.2	3.2	420	1	US-08-470-179-148	Sequence 148, App
31	39.2	3.2	2353	6	PCT-US92-06840-1	Sequence 1, Appl
32	39.2	3.2	28804	3	US-08-592-874-1	Sequence 1, Appl
33	39	3.2	28804	5	US-09-096-942-2	Sequence 2, Appl
34	39	3.2	43280	3	US-08-804-227C-1	Sequence 1, Appl
35	38.8	3.1	8051	4	US-08-576-626A-2	Sequence 2, Appl
36	38.8	3.1	11219	2	US-07-642-734C-1	Sequence 1, Appl
37	38.6	3.1	1187	2	US-08-440-856A-2	Sequence 2, Appl
38	38.6	3.1	3231	2	US-08-074-121-4	Sequence 4, Appl
39	38.6	3.1	3231	6	PCT-US94-06447-4	Sequence 1, Appl
40	38.4	3.1	1215	5	US-08-947-726A-1	Sequence 1, Appl
41	38.4	3.1	2048	5	US-08-776-251-1	Sequence 20, Appl
42	38.2	3.1	2109	4	US-08-555-568B-20	Sequence 1, Appl
43	38	3.1	30001	1	US-08-125-468-1	Sequence 1, Appl
44	38	3.1	30001	3	US-08-474-933-1	Sequence 1, Appl
45	37.8	3.1	2588	3	US-08-796-414B-6	Sequence 6, Appl

## ALIGNMENTS

RESULT 1  
US-08-816-105A-2  
Sequence 2, Application US/08816105A  
Patent No. 5989882  
GENERAL INFORMATION:  
APPLICANT: Gruenger, Anneliese; Dellweg, Hans-Georg; Lenz,  
APPLICANT: Jurgen; Schroeder, Werner; Pape, Hermann;  
APPLICANT: Goekle, Klaus; Schaper, Beate; Hemker, Michael;  
APPLICANT: Plepersberg, Wolfgang; Distler, Jurgen;  
APPLICANT: Stralmann, Ansgar  
TITLE OF INVENTION: PROCESSES FOR PREPARING ACARYOSYL  
TITLE OF INVENTION: TRANSFERASE AND FOR USING IT IN THE  
TITLE OF INVENTION: CONVERSION OF ACAROSE HOMOLOGUES  
TITLE OF INVENTION: INFO ACAROSE, FOR THE PREPARATION  
TITLE OF INVENTION: OF ACAROSE HOMOLOGUES  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Sprung Kramer Schaefer & Briscoe  
STREET: 660 White Plains Road  
CITY: Tarrytown  
STATE: New York  
COUNTRY: USA  
ZIP: 10591-5144  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 MB storage  
COMPUTER: Apple Macintosh 6500  
OPERATING SYSTEM: System 7.5  
SOFTWARE: WordPerfect 3.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816, 105A  
FILING DATE: 14-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 19625269.5  
FILING DATE: 25-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 19611252.4  
FILING DATE: 22-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurt G. Briscoe  
REGISTRATION NUMBER: 33, 141  
REFERENCE/DOCKET NUMBER: Bayer 9814-KGB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (914) 332-1700  
TELEFAX: (914) 332-1844  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2582 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: double

TOPOLOGY: linear  
US-08-816-105A-2

Query Match 3.7%; Score 46; DB 4; Length 2582;  
Best Local Similarity 47.3%; Pred. No. 0.013;  
Matches 139; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

352 CCGCGTCACTGCTTGCACCGATGCGTTGTGATTCGGGACACCAATGCGCTACG 411  
1576 CCGCGTCTGCGCGGCGACCTTACCGACGCTGCGCGGCGCTTCCGGGCTCGAC 1635  
412 AGACGTCACAGCCCGCGACGCGACCTACCTGCTTCCCGGATGACGAGTCCGCG 471  
1636 TGACCGTGAGAGACGCGGACGCGACCGACCGCTCGCGCGGCTGCGGCGCG 1695  
472 GTGCACTGCGCGCGCGGCTTACCGCTGATGCGAACCCAGGGGCGCGCGGT 531  
1696 GCGAGGTGCGCGCTGCTGCTGCTACCGCGCGCGCGGACCGACCGCGGCGG 1755  
532 TCGTATGCGCGGCGGCGCGCGCGGCGCGGCGCGGCGGCGGCGGCGGCGG 591  
1756 TCGGCGCGGCTGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1815  
592 CCGCGCGCTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 645  
1816 CCGCGCGGACCGCTGCGGATCGCGCGGCGCGCGCGCGCGCGCGCGCGCG 1869

## RESULT 2

Sequence 3, Application US/08403852D

Patent No. 5891695

GENERAL INFORMATION:

APPLICANT: Blanc, Veronique  
APPLICANT: Crouzet, Joel  
APPLICANT: Jacques, Nathalie  
APPLICANT: Lacroix, Patricia  
APPLICANT: Thibaut, Denis  
APPLICANT: Zagorec, Monique  
APPLICANT: Debussche, Laurent  
APPLICANT: De Crecy-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved In The  
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/08/403,852D

FILING DATE: 10-MAY-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR 93/00923

FILING DATE: 25-SEP-1993

PRIOR APPLICATION DATA: FR 92/11441

FILING DATE: 25-SEP-1992

ATTORNEY/AGENT INFORMATION:

NAME: Meyers, Kenneth J.

REGISTRATION NUMBER: 25,146

REFERENCE/DOCKET NUMBER: 03806, 0054-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 833 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: S.pristineaspiralis

FEATURE:

NAME/KEY: CDS

LOCATION: 1..833

US-08-403-852D-3

Query Match 3.7%; Score 45.4; DB 3; Length 833;  
Best Local Similarity 48.0%; Pred. No. 0.014;  
Matches 130; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

542 CCGGCG 601  
327 CCGGCG 386  
602 CAAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 661  
387 CAGCG 446  
662 CGACAACTTGGCGCACTGCGAGCGCGCGCGCGCGCGCGCGCG 721  
447 ACTGACACACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 506  
722 GCGTACAGAGCTCGAGGCTGCGCGCGCGCGCGCGCGCGCGCG 781  
507 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 566  
782 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 812  
567 CCGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 597

## RESULT 3

Sequence 1, Application US/08403852D

Patent No. 5891695

GENERAL INFORMATION:

APPLICANT: Blanc, Veronique  
APPLICANT: Crouzet, Joel  
APPLICANT: Jacques, Nathalie  
APPLICANT: Lacroix, Patricia  
APPLICANT: Thibaut, Denis  
APPLICANT: Zagorec, Monique  
APPLICANT: Debussche, Laurent  
APPLICANT: De Crecy-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved In The  
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852D
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5392 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
US-08-403-852D-1

```

```

Query Match          3.7%; Score 45.4; DB 3; Length 5392;
Best Local Similarity 48.0%; Pred. No. 0.023;
Matches 130; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 542 CCGGGTCCCGGGCTGCAACCGCCGACGTCGTGATGCGCGCGGCAACCGCGGCTA 601
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3018 CGAGGGCGCGACCGGCACTCATCGACGTCGCCGCGCTGTGGACAGCTCGACGA 3077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 602 CACGAGCGCGGATCGCAAGGCGGACCGGACCGGTTACGGTTCTACATCAACAT 661
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3078 CGAGCGCTTGTCTACACAGCCGCGCGGCTGTACTGCGGGTGTCCCGCTCCACCA 3137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 662 CGACAACTTGGCACTCGACCGCGAGTTCTGCGCGCGGATCCACACTCGTACTATC 721
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3138 ACTGACACCAAGGAGGAGCACTTCAGACGTGGCGCGGCCCCCTCAAGCTGCCCGCGCC 3197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 722 GGCCTACGAGCTGAGAGGTCCTGCAACGTCGCCGACCTGCTGATTGGGCGCTCTGT 781
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3198 GCAGGGCGACCCCGTCTTCGCGCTCACCGGCGCCCGCTCGCGCGCGCGACCTCGT 3257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 782 GCCAGGCGCAAGGACCCCAATTAGTCTCG 812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3258 CTTGCTGACAGAGCGCGCGACCGCGCTCG 3288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 4
US-08-258-261B-6
Sequence 6, Application US/08258261B
Patent No. 5639949
GENERAL INFORMATION:
APPLICANT: Schnupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
APPLICANT: Uknes, Scott Joseph
TITLE OF INVENTION: Genes for the synthesis of
antipathogenic substances
NUMBER OF SEQUENCES: 22

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,261B
FILING DATE: 08-JUN-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-258-261B-6

```

```

Query Match          3.5%; Score 43.6; DB 1; Length 28958;
Best Local Similarity 44.8%; Pred. No. 0.098;
Matches 290; Conservative 0; Mismatches 354; Indels 4; Gaps 3;

QY 148 GGCATGAGGTGCTCTTCCAGCACTGCGGAGAGGCGCTCGGCTATCCAGCAGCGGAT 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25306 GCCCAGAGAGACTTACCTCGCGGATGTGCCCTATTGCTGCGCACACCGCGCCAC 25365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 208 TTCAAGCGCGACGCGCGCACTGTGCGCACCGCGGACGAGTGTGGCGGACGCTGAT 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25366 TTGAGACACCGCGCGCTCTGTGAGCCCAACCGGAGAGAGCTCTCTCGGGCTCGAC 25425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 268 TTATGCTCAAGTCAA-AGAACCGATAGCGGCGGAATACGGCGCGCTCGACAGGGCG 326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25426 TCGCTGCCAGACAGACCGCGCGCGGACCGCTCTCGGAGGAGGAGGAGCCAGCGGC 25485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 327 ATCTGTGACGTTCTTGATTTGGCGGCTGACGCTTGACCGATCGTTGTGAT 386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25486 AAGCTGCTTCTGCTCTTCTTGCGCAAGGCTCGCACTGGGAAGGATGGCCCTCTCGTG 25545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 387 TCGGACCACTGCAATTGCTACGAGACGCTCGACAGCCGCGGCGGCACTACCCCTG 446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25546 CTGCACTCTCGCGCTCTTCCGCTCAGCTCGAAGCATGCGAGCGGCTCGCTCT 25605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 447 CTGCGCCGATGAGCAAGTCCGCGGTGCGACGCGCGCGCGGCTGGGCTTACCAAC--C 504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25606 CAGCTGAGTGAAGCTGCTGCTGCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 25665
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 505 TGATGGAACCCAGGAGGCGCGGCTGCTGATGCGCGGCGGCTGCGGAGCGG 564
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25666 CGGCTGAGCTGTACAGCCCGCGCTCTTGGCGTATGCTCTCTGCGGCGCTCTG 25725
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 565 CCGAGC-TGCTGTGATGCGCGCGCGGCGGCGGCTTACAGCAGCAGCGCATCGCAAC 623
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25726 CGCTCGTGGGTAGAGCCCGCGCGCTGCTGCGCACAGTCAGGGGCGAGATCGCGCC 25785
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



QY	634	GGCAATGGGGGCACCGTATACGGTTTACGATCAATCAATGACAATTTGGCAATCGAC	683
Db	25786	GCCTTCGTGCGAGGGCTCTCTCCCTCGAGAGACGGGGCCGCATCGCCGCCCTGGCGAAC	25845
QY	684	GCCGAGTTCTGGGGCGGATCCACACTGCTACTCATGGGCTTAGAGTCAAGSTGGC	743
Db	25846	AAAGGCTTCACACCGTGGCGGCAACGGGGCGATAGCGCGCTGTGAGCTCGGGCGCTCC	25905
QY	744	GTCAAACGTGGCAGCTGTATTTGGGGCGTCTGTGTCACAGGGGCC	791
Db	25906	GACCTTCACACTTACTCTGCTCTCTGGGGGCAAGGCTTCCATCGCC	25953

RESULT 5

Sequence 6, Application US/08456837  
Patent No. 5643774  
GENERAL INFORMATION:  
APPLICANT: Schnupp, Thomas  
APPLICANT: Ligon, James M.  
APPLICANT: Beck, James Joseph  
APPLICANT: Hill, Dwight Steven  
APPLICANT: Ryals, John Andrew  
APPLICANT: Gaffney, Thomas Deane  
APPLICANT: Lam, Stephen Ting  
APPLICANT: Hammer, Phillip E.  
APPLICANT: Urnes, Scott Joseph  
TITLE OF INVENTION: Genes for the synthesis of  
antipathogenic substances  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:

STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,837  
FILING DATE: 01-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/457,205  
FILING DATE: 01-JUN-1995  
APPLICATION NUMBER: 08/258,261  
FILING DATE: 08-Jun-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28958 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
OS-08-456-837-6

Query Match	3.5%;	Score 43.6;	DB 1;	Length 28958;
Best Local Similarity	44.8%;	Pred. No. 0.098;		

Matches	290;	Conservative	0;	Mismatches	354;	Indels	4;	Gaps	3;
OY	148	GGCCATGAGTCTCTATCCAGCAGTGC	CGGAGAGGGCTCGGCTATACCGACGGGAT	207					
Db	25306	GGCCAGCAGACCTTACCTCGGGATG	GGGCTTATCGGTCCGCCACACCGGCCAC	25365					
OY	208	TTCAAGGGGGCAGGCGCCAACTGGT	CGGCACCGCCGACAGGTGTGGGCGCAGCTGAT	267					
Db	25366	TTCCAGCAACCGGGCGCTTCGTACC	CCCAACCGCGACAGACTCTCTCGCGCTCGAC	25425					
OY	268	TTATTGCTCAGGCTAA-AGAACGAT	ATACGGGGAATACGGCGCCTTCGCACAGGGCG	326					
Db	25426	TCGCTCGCCCAAGCAACCGCCCGG	AGACCTCTCTCGACGAGGGGAAGCCACGGC	25485					
OY	327	ATCTTGTCACGCTCTTGATTTGGC	CGCGTACGCTGTGCACGATGCGTTGTGAT	386					
Db	25486	AAGCTGCTTCTGCTTTTCTTGGGC	AAAGGCTTCGAGTGGGAAGGATGCGCTTCTGCTG	25545					
OY	387	TCGGCACCACGCTCAATTGCTACG	AGACCGTCCAGACCGCGGCGACTACCCCTG	446					
Db	25546	CTCGACTCTCTCCCGCTTTCGGC	CTCAGCTCGAAGCATGGAGCGCGCGCTCTGCTCT	25605					
OY	447	CTTGCCCGGATGAGGAAGTCCCGG	TGACTCGCGCCGACGTTGGGGCTTACAC--C	504					
Db	25606	CACGTGAGTGGAGCTGCTGCGCTC	TCTGCGCGCGACGAGGGCGCCCTCTCGAC	25665					
OY	505	TGATCGAACCCMAAGGGGGCCGG	TGTGCTGATGAGGGGGGTGCGCGGCTCGAAACGG	564					
Db	25666	CGCGTCGACGTGTGTACAGCCGCC	CTTTGGCGTATGTTCTCCCTGGCGCCCTCTGG	25725					
OY	565	CCGAGC-TCGTGTATGGGGCGGG	CGGACCGCGGCTTACACGCAAGCCGCTATGCCAAC	623					
Db	25726	CGGTGCTGGGGTATGAGACCCGCC	CCCTGTGCGGCACAGTCAAGGGGAGTATCCCGCC	25785					
OY	624	GGCATGGGCGCACCGTTACAGCTT	TTAACAATCATGTGACATAAATTTCGGAACTCGAC	683					
Db	25786	GGCTTGTGTGAGGGGCTCTCTCC	TTCAGAGACGGGGCCCGATGCGCGCCTTGGCGCAC	25845					
OY	684	GCCGAGTTCGTGGGCGCGGATCC	ACACTGCTACTATCGGCCCTACGAGCTCGAGGGTGC	743					
Db	25846	AAAGGCTCACACCGTATGCGCGG	CAACGGGGCAATGAGCGCGCTGTAGAGTGGGCGCTTC	25905					
OY	744	GTCAAACGTCGGACCTGGTGATT	TGGGGCGTCTGTGTGTCACAGGGGC	791					
Db	25906	GACCTTCAGACTTACTCTGCTCTG	GGGCAACAGGCTCTCTCATGGC	25953					

## RESULT

1 US-08-457-342-6  
 2 Sequence 6, Application US/08457342  
 3 Patent No. 5662898  
 4  
 5 GENERAL INFORMATION:  
 6 APPLICANT: Schupp, Thomas  
 7 APPLICANT: Ligon, James M.  
 8 APPLICANT: Beck, James Joseph  
 9 APPLICANT: Hill, Dwight Steven  
 10 APPLICANT: Ryals, John Andrew  
 11 APPLICANT: Gaffney, Thomas Deane  
 12 APPLICANT: Lam, Stephen Ting  
 13 APPLICANT: Hammer, Phillip E.  
 14 APPLICANT: Uknes, Scott Joseph  
 15  
 16 TITLE OF INVENTION: Genes for the synthesis of  
 17  
 18 TITLE OF INVENTION: antipathogenic substances  
 19  
 20 NUMBER OF SEQUENCES: 22  
 21  
 22 CORRESPONDENCE ADDRESS:  
 23 ADDRESSEE: Ciba-Geigy Corporation  
 24 STREET: 7 Skyline Drive  
 25 CITY: Hawthorne  
 26 STATE: NY  
 27 COUNTRY: USA  
 28 ZIP: 10532  
 29  
 30 COMPUTER READABLE FORM:  
 31 MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/457,342  
 FILING DATE: 01-JUN-1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/457,205  
 FILING DATE: 01-JUN-1995  
 APPLICATION NUMBER: 08/258,261  
 FILING DATE: 08-Jun-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Elmer, James Scott  
 REGISTRATION NUMBER: 36,129  
 REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-541-8614  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 28958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-457-342-6

Query Match 3.5%; Score 43.6; DB 1; Length 28958;  
 Best Local Similarity 44.8%; Pred. No. 0.098;  
 Matches 290; Conservative 0; Mismatches 354; Indels 4; Gaps 3;

148 GCCCATGAGTCTCATCAGGAGGCGGAGAGGCGTCCGTATCAGCCAGCGGAT 207  
 25306 GCCCAGCGACCTCCACCGGATGTGGCTATTCCTGCTACCGCCGCGCCAC 25365  
 208 TTCAAGGGGAGGCGGCACTGGTCGGCAGCCGCAAGGTGTGGCGAGCTGAT 267  
 25366 TTCGAGCAGCGCGCGCTCTGTAGCCACACCGCAGAGCTCTCCGGCTGAC 25425  
 268 TTATTCCTAAGTCAA-AGAACCAGATACGGCGAAATACGGCCCTCGGACAGCGGCG 326  
 25426 TCGCTCGCCAGGAGAACCGCCCGCCAGCAGCTCTCGGAGCAGCGAGGACGAGCG 25485  
 327 ATCTTTTCACTGTTGATTTGGCCCGCTCAGTGTGTCACGATGGTGTGAT 386  
 25486 AAGTCGCTTCTCTTCTGCGGCAAGCTGCGAGTGGAGGAGTGGCTCTCCGCTG 25545  
 387 TCCGCGACGATCAATGCTTACAGAGACCGTCCAGACCGCGAGCGGCACTACCGCTG 446  
 25546 CTCGACTCTGCGCGCTTTCGGGCTCAGTCAAGCATCGAGCGCGGCTC 3TCTCT 25605  
 447 CTGCCCCGATGAGGAGTCCCGGCTGACTCCGCCGCCAGTTGGCGCTTACAC--C 504  
 25606 CACGTCGAGTGAAGCTGCTGCTGCGCGCGAGCA GCGGCTCTCTCTCTGAC 25665  
 505 TGATCGGAACCAAGGGGCGCGGTGCTGATGGGCGGCTCCGCGCTCGAACCG 564  
 25666 CGCGTCGAGCTGTCACACCGCCCTCTTTGCCGTCAGTCTCTCCGCGCGCTCTGG 25725  
 565 CCGAGC-TCGTGTGATGCGCGCGGACCGCGGCTACAGACGAGCCCGATGCCAAC 623  
 25726 CGCTGCTGCGGCTGAGACCGCGCGCTGTGCGGCGACATCAGGGGAGATGCCGCC 25785  
 624 GGCATGGCGGAGCGGATGAGTTCTAGACATCAACATCGAACAACTCGCAACTCGAC 683  
 25786 GCTTCGTCGAGGCGCTCTCTCTCTCGAGAGCGGCGCGCATTCGCCCTCGCGACG 25845  
 684 GCGAGTCTGCGGCGGATCCACACTGCTACTGATGCGCTACGAGC CGAGGGTGGC 743

Db 25846 AAAGCGCTACACCGCTGCGCGGCAAGGGGCGATGCGCCCGCTCGAGCTCGCGCTCC 25905  
 Oy 744 GTCAAAGCTGCGGACCTGCTGATGGGCGCGCTCTGTGCGCAGCGCC 791  
 Db 25906 GACCTCCAGACTTACTCTCTCCCTGGGCGCAGAGCTCTCCATCGCC 25953

RESULT 7  
 US-08-457-646A-6  
 Sequence 6, Application US/08457645A  
 Patent No. 5679560

GENERAL INFORMATION:  
 APPLICANT: Schnupp, Thomas  
 APPLICANT: Ligon, James M.  
 APPLICANT: Beck, James Joseph  
 APPLICANT: Hill, Dwight Steven  
 APPLICANT: Ryals, John Andrew  
 APPLICANT: Gaffney, Thomas Deane  
 APPLICANT: Lam, Stephen Ting  
 APPLICANT: Hammer, Phillip E.  
 APPLICANT: Uknes, Scott Joseph  
 TITLE OF INVENTION: Genes for the synthesis of  
 TITLE OF INVENTION: antipathogenic substances  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ciba-Geigy Corporation  
 STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: NY  
 COUNTRY: USA

ZIP: 10532  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-/S/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/457,646A  
 FILING DATE: 01-JUN-1995  
 CLASSIFICATION: 530

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/457,205  
 FILING DATE: 01-JUN-1995  
 APPLICATION NUMBER: 08/258,261  
 FILING DATE: 08-Jun-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Elmer, James Scott  
 REGISTRATION NUMBER: 36,129  
 REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-541-8614  
 TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 28958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-457-646A-6

Query Match 3.5%; Score 43.6; DB 1; Length 28958;  
 Best Local Similarity 44.8%; Pred. No. 0.098;  
 Matches 290; Conservative 0; Mismatches 354; Indels 4; Gaps 3;

148 GCCCATGAGTCTCATCAGGAGGCGGAGAGGCGTCCGTATCAGCCAGCGGAT 207  
 25306 GCCCAGCAGACTTCACTCTCGGAGATGTGGCTATTCCTGCGCAGCAGCGCGCCAC 25365  
 208 TTCAAGCGGAGCGCGCACTGTGCTGCGACCGCCGAGGAGTGTGGCGCAGCTGAT 267

```

Db 25366 TTCGAGACCGGCGCCCTCTAGCCCAACCGGAGAGACTCTCTCCCGCTCGAC 25425
Oy 268 TTATGCTCAAGTCAA-AAGAACGATACGGCGGATACGGCCCTCGACACGGCG 326
Db 25426 TCGCTCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 25485
Oy 327 ATCTGTTCACTTCTTGATTTGGCCGGGTGACCGCTTGACCGAGATGCTTGTGAT 386
Db 25486 AAGCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25545
Oy 387 TCCGACACCACTCAATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 446
Db 25546 CTCGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25605
Oy 447 CTTCGCGCGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 504
Db 25606 CACGTCGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 25665
Oy 505 TGATGGAGAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 564
Db 25666 CCGCTGAGAGTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 25725
Oy 565 CCGAGC-TCGTGTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 623
Db 25726 CGCTGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 25785
Oy 624 GCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 683
Db 25786 GCGTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 25845
Oy 684 GCGGAGTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 743
Db 25846 AAGCGCTACCAACCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 25905
Oy 744 GTCAAACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 791
Db 25906 GACCTCCAGACTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25953

```

## RESULT 8

US-08-458-076A-6

Sequence 6, Application US/08458076A

Patent No. 5698425

## GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Lyon, James M.

APPLICANT: Beck, James Joseph

APPLICANT: Hill, Dwight Steven

APPLICANT: Ryals, John Andrew

APPLICANT: Gafney, Thomas Deane

APPLICANT: Lam, Stephen Ting

APPLICANT: Hammer, Phillip E.

APPLICANT: Unnes, Scott Joseph

TITLE OF INVENTION: Genes for the synthesis of

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESS: Ciba-Geigy Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08-458-076A

FILING DATE: 01-JUN-1995

CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/457, 205

FILING DATE: 01-JUN-1995

APPLICATION NUMBER: 08/258, 261

FILING DATE: 08-JUN-1994

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129

REFERENCE/DOCKET NUMBER: CGC 1506/CIP3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8614

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 28958 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-458-076A-6

## Query Match

Best Local Similarity 44.8%; Pred. NO. 0.098;

Matches 290; Conservative 0; Mismatches 354; Indels 4; Gaps 3;

```

Oy 148 GCCATGAGTGTCTATCAGGAGGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 207
Db 25306 GCCACGACGACCTACACCTCGGAGATGTGCTATTGCTGGACACAGCGGCGCAC 25365
Oy 208 TTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 267
Db 25366 TTGAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 25425
Oy 268 TTATGCTCAAGTCAA-AAGAACGATACGGCGGATACGGCGGCGGCGGCGGCGGCGG 326
Db 25426 TCGCTCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 25485
Oy 327 ATCTGTTCACTTCTTGATTTGGCCGGGTGACCGCTTGACCGAGATGCTTGTGAT 386
Db 25486 AAGCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25545
Oy 387 TCCGACACCACTCAATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 446
Db 25546 CTCGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25605
Oy 447 CTTCGCGCGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 504
Db 25606 CACGTCGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 25665
Oy 505 TGATGGAGAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 564
Db 25666 CCGCTGAGAGTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 25725
Oy 565 CCGAGC-TCGTGTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 623
Db 25726 CGCTGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 25785
Oy 624 GCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 683
Db 25786 GCGTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 25845
Oy 684 GCGGAGTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 743
Db 25846 AAGCGCTACCAACCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 25905
Oy 744 GTCAAACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 791
Db 25906 GACCTCCAGACTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25953

```

RESULT 9  
 US-08-764-233A-4  
 Sequence 4, Application US/08764233A  
 Patent No. 5716849  
 GENERAL INFORMATION:  
 APPLICANT: Ligon, James M.  
 APPLICANT: Schupp, Thomas  
 APPLICANT: Beck, James J.  
 APPLICANT: Hill, Dwight S.  
 APPLICANT: Neff, Snezana  
 APPLICANT: Ryals, John A.  
 TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ciba-Geigy Corporation  
 STREET: 520 White Plains Road, P.O. Box 2005  
 CITY: Tarrytown  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10591  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/764,233A  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/729,214  
 FILING DATE: 09-OCT-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/258,261  
 FILING DATE: 08-JUN-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meigs, J. Timothy  
 REGISTRATION NUMBER: 38,241  
 REFERENCE/DOCKET NUMBER: 1506/CIP6  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (919) 541-8587  
 TELEFAX: (919) 541-8689  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 28958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 ORIGINAL SOURCE:  
 ORGANISM: Sorangium cellulosum  
 IMMEDIATE SOURCE:  
 CLONE: P98/1  
 US-08-764-233A-4

Query Match	Similarity	Score	DB 2	length
Best Local	44.8%	No. 0.098		28958
Matches	290	Conservative	0	Mismatches 351; Indels 4; Gaps 3
QY	148	GGCCATGAGTGTCTCATCCAGCAGGTGCCGCGAGAGGCTCGGCTATCAACCGACGGGAT	207	
Db	25306	GGCCACGACAGCACTTCCCTCCGCGATGTGGCTATTCTGCTGGCCACACCCCGCCAC	25365	
QY	208	TTCAAGGGGGGACGGCGGCACTGCTGGGACCGCCGACAGGTGTGGGCGGACGCTGAT	267	
Db	25366	TTTCGAGCACCGCGCGCTTTCGTAGCCACAACCGCAGAGAGTCTCTTCGCGGCTCGAC	25425	
QY	268	TTATTGCTCAAGGTCAA-AGAACCGATTACGGCGGCAATAGCGCGCGCTCGACACGGGAG	326	
Db	25426	TGGCTGTGCCCAAGACAAAGCCGCGCCGACACACGTCTCTGGAGAGGAGCGGAACCCACGGG	25485	
QY	327	ATCTTGTTACGCTTCTTGCAATTTGGCCGCGTCACTGCTTT-ACCGATGCGTGTGGAT	386	

Db	25486	AACCTGCTTGGTCTTCTTCTTCTGGGCAAGCGCTCGCAAGTGGGAAGGATGGCCCTCTCGTG	254545
QY	387	TCCGGCACCACGTCGAATTGECCTACGAGACGCTCCAGACCCGCGAGCGACCTAACCCCTG	446
Db	25546	CTCGACTCTCCGCCCGCTCTTCGGCGCTCAGCTCGAAGCATCGAGCATCGGCGCTCGCTCT	256055
QY	447	CTTGCCCCCATGAGCAAGTGCGCCGCTGACATCGCCGCCCCAGGTTGGCGCTTACAC	-C 504
Db	25606	CACGTGAGTGGAGCGCTGCTCCCTCCCTCGCGCGCGACGAGGGCGCCCTCTCCCTCGAC	256655
QY	505	TGATGGAAACCCAAAGGGGGCGCGGTGTGCTGATGGGGGGGTGCCCGCGTCGAACCG	564
Db	25666	CGGCTGACAGCTGCTACAGCGCCCGCCTCTTTGCCGTATGGTCTCTCTGGGGCCCTCTGG	257255
QY	565	CCGACG-TGGTGTGATCGGCGCGGACCGCGCGCTTCAACGCGACGCCCATCGCCAC	623
Db	25726	CGCTCGCTGGGGGTAGAGCGCCCGCCCTCTGTGGCCACATCAGGGGAGATCGCGCGC	257855
QY	624	GGCATGGGCGCGACCGTTAGGTTCTAGACATCAACATCGACAAATCTGGCAACTCGAC	683
Db	25786	GCCTTCGTGAGGGCGCTCTCTCCCTCGAAGACGGGGCGCCCATCGCCCGCTCGGCACG	258455
QY	684	GCGGAGTTCTGGGGCGGATCCACACTGCTACTCATCGGCTTACGAGCTCGAGGGTGGC	743
Db	25846	AAAGCGCTCACACACCGCTCCCGCGCAACGGGGCATATGGCGCGCGTCGAGCTCGGCGCTCC	259055
QY	744	GTCAAACGTGCGACCTGGTGAATTGGGGCGTCCCTGGGTGCCAGGCGCGC	791
Db	25906	GACCTCGAGACTTACTCGCTCTCTGGGGGCGACAGGCTTCTCATCGCC	25993

RESULT 10  
 US-08-457-335A-6  
 Sequence 6, Application US/08457335A  
 Patent No. 5723759  
 GENERAL INFORMATION:  
 APPLICANT: Schupp, Thomas  
 APPLICANT: Ligon, James M.  
 APPLICANT: Beck, James Joseph  
 APPLICANT: Hill, Dwight Steven  
 APPLICANT: Ryals, John Andrew  
 APPLICANT: Gaffney, Thomas Deane  
 APPLICANT: Lam, Stephen Ting  
 APPLICANT: Hammer, Phillip E.  
 APPLICANT: Uknes, Scott Joseph  
 TITLE OF INVENTION: Genes for the synthesis of  
 TITLE OF INVENTION: antipathogenic substances  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ciba-Geigy Corporation  
 STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10532  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentln Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/457,335A  
 FILING DATE: 01-JUN-1995  
 CLASSIFICATION: 860  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/457,205  
 FILING DATE: 01-JUN-1995  
 APPLICATION NUMBER: 08/258,261  
 FILING DATE: 08-Jun-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Elmer, James S.,olt  
 REGISTRATION NUMBER: 36,129

```

1 REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
2 TELECOMMUNICATION INFORMATION:
3 TELEPHONE: 919-541-8614
4
5 TELERFAX: 919-541-8689
6
7 INFORMATION FOR SEQ ID NO: 6:
8
9 SEQUENCE CHARACTERISTICS:
10
11 LENGTH: 28958 base pairs
12
13 TYPE: nucleic acid
14
15 STRANDEDNESS: single
16
17 TOPOLOGY: linear
18
19 MOLECULE TYPE: DNA (genomic)
20
21 HYPOTHETICAL: NO
22
23 ANTI-SENSE: NO
24
25 US-08-457-335A-6

```

Query Match	3.5%;	Score 43.6;	DB 2;	Length 28958;
Best Local Similarity	44.8%;	Pred. No. 0.098;		
Matches 290;	Conservative 0;	Mismatches 35%;	Indels 4;	Gaps 3

Oy	148	GGCATTGAGAGTCTCATTCACAGGACAGGATGGCGAGAGAGGCTCGTCAATACCAACGGCGAT	207
Db	25306	GGCCAGAGAGACTCAACCTTCGCGGATGAGGCTATTGCGTTATGCGTGGCCACACACCGCGGCCAC	25365
Oy	208	TTCAAGGCGGACAGGCGGCGCACTGGTCGGACACCGCGGACACAGTGTGGGCGCACGCTAT	267
Db	25366	TTTCGAGACACCGGCGGCTCTTCGTACCCCAACACCGGAGAGACTCTCTTCGCGCTCGAC	25425
Oy	268	TTATTGCTCAAGCTCA--AGAACCGATNAGCGGGAATACGCGCGCGTCGACACGGGCG	326
Db	25426	TGCGTCGCCACGAGCAAGCCCCCGCGAGACCGTCTTCGACGAGGAGCGAAGCCACGGC	25485
Oy	327	ATCTGTTCACCTTCCTTGATTTGGCGCGGTACAGTGTGACCGAGTCCGTTGTGGAT	386
Db	25486	AAGCTGTGTTGTTCTTCTTCGTGGGCAAGGCTCGAGTGGG--AGGATGACCTTCCTCGTG	25545
Oy	387	TCCGGACACACGTCAATTGCTTACGAGACCGTCCAGACCGCGGACGCGCATACCCCTG	446
Db	25546	CTCGACTCCCTCCCGCTTCCGCGCTCAGCTCGAAGCATGCGACGCGCGCTCCCTCT	25605
Oy	447	CTTGGCCCCATGAGCAATTCGCGCGTCACTTCGCGCGCCAGATTGGGCTTACAC--C	504
Db	25606	CACGTGAGATGAGACTGCTCGCCGCTCCGCGCGCGAGAGAGGGGCGCCCTCCCTCGAC	25665
Oy	505	TCATCGAACCACAAGGGGCGCGGGTGTGCTATGGGCGGGGTGCCCGCGCTGCAACGG	564
Db	25666	CCGCTGACGTGTATACAGCCGCCCTTGTGCGTACGTATGTTCCCTGGCGCCCTGTGG	25725
Oy	565	CCGACG--TCGTGTGATCGGCGCGCGACCGCGGCTTACACAGCAGCCCGCATCCGCAC	623
Db	25726	CGCTCGCTGGGGTANAGCCGCGCGCGTCTGTGGGCAACATGACGCGCAGATTCGCGCG	25785
Oy	624	GGCATGGGCGGACGCTTACGGTTCTAGACATCAACATGACACAACTTGGGCACTCGAC	683
Db	25786	GGCTTGTGTGCAAGGCCCTCTCTCCCTCGAGGAGCGCGCCGATCGCGCCCTGGCGAC	25845
Oy	684	GCCGATTCGTGGGCGCGGATTCACACTCCACTACTCACTCGGCTACGATCAGAGGTGGC	743
Db	25846	AAAGCGCTACACACGCTCCCGGCAACGGGGCCATGGCCGCGTGAAGCTCGGCGGCTCC	25905
Oy	744	GTCAAACGTGCGGACTGTGATTTGGGGCGCTCTGTGTGCAAGCGCC	791
Db	25906	GACCTTCAACCTTACCTGCTCCTTGGGCGAGAGGCTCTTCATATGCC	25953

RESULT 11  
US-08-729-214-6  
Sequence 6, Application US/0872922A  
Patent No. 5817502  
GENERAL INFORMATION:  
APPLICANT: Ligon, James M.  
APPLICANT: Hill, Dwight Steven  
APPLICANT: Ryals, John Andrew

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

Query Match	3.58;	Score 43.6;	DB 2;	Length 28958;
Best Local Similarity	44.88;	Pred. NO. 0.098;		
Matches 290;	Conservative	0;	Mismatches 354;	Indels 4;
				Gaps 3;

OY	148	GGCATTAGAGTGGCTCTTCCAGGCAAGTCCCGAGAGGGCTGGCTATACACCGAGGGAT	207
Db	25306	GCCACGACGACACTCAACCTCGCGGATGTGGCTATTGGCTGGGCACACCGCGGCCAC	25365
OY	208	TTCAAGCGGAGGACGCGCGACACTGGTGGCAGCGCCGACACAGGTGGGGCCAGCGCTGAT	267
Db	25366	TTTCAGACCGCGGCGGCTCTCTGTACCCACAAACCGCAGCGAGCTCTCTCGGCGTCGAC	25425
OY	268	TTATTTGCTCAAGTCAA - AGAACCCATATAGCGGCGGAATACGGCCGCTGGGACACGGGCG	326
Db	25426	TCGCTCGCCCGAGACAAAGCCGCCGCCGAGACCGCTCTCGGACGAGACCGGAACCCAGCGC	25485
OY	327	ATCTTTGTACGTTCTTCGATTGTGGCCCGACAGCGTGGTGTCAACCGATGGTGGTGTGGAT	386
Db	25486	AAGCTGCTCTTGGCTCTTTTCTTGCGCAAGGCTGCGAGTGGGAAGGATGGCCCTCTTGCTG	25545
OY	387	TCCGGCACACAGTCATTGCTCTACGAGACCGTCCAGACCGCGCGACGACATACCCCTG	446
Db	25546	CTCGACTCTCTGGCCGCTCTTCCGGCGGTCAAGTCCGAGAGATCGAGCGCGGCTCGCTCTCT	25605
OY	447	CTTCCGCCCATGAGCCAAAGTCGCGCGGTGCACTCGCCGCCACAGTTGGCGCTTACCAAC - C	504
Db	25606	CACGTCAGAGGAGCCCTGGCTCGCGCTCGCTCGCGCCGCGACGAGGGCGGCCCTCTCCCTGAC	25665
OY	505	TGATGCGAACCCAAAGGGGGCGGGGTGCTGATGATGGCGGGGTGCCCGGGTGTGGAAACCGG	564
Db	25666	CGGCTCTACGATCAGAGCCCGCCCTTTTGCGGTGATGATGCTCTCCCTGGGCGGCGCTCTGG	25725

```

0Y      565  CCGAGC-TGCTGTATCGGCGCGGACACCGCGGCTAACAGCAGCCGCATCGCCAC 623
Db 25726 CGCTGCTCGGCGGTAGAGCCCGCCCGCGCTGCGGCCACAGTCAGGCGCAGATCGCCGC 25735
0Y      624  GGCATGGGCGCACCTTACGGTTCTAGACATCAACATCGACAACTTGGCACTGCAC 683
Db 25786 GCGTTCGTCTCGAGGCGCTCTCTCCCTCGAGAGACGGGCGCATGCGCCGCTGCGCAGC 25845
0Y      684  GCGGATCTTCGGGCGCGATCCACACTGCTCTACTCATGGCCCTACGAGGTGCAGGTGTC 743
Db 25846 AAGGCGCTACACCGTCGCGCCGCAACGGGCGCATGCGCGCGTGCAGGTGCGCGCTCC 25905
0Y      744  GTCAAACGTGCGCACCTGTGATGTGGGCGCGCTCTGTGTCCAGCGCGC 791
Db 25906 GACCTTCAGACCTACCTCGCTCCCTCGGGGCGACAGGCTCTCATGCGCC 25953

RESULT 12
: US-08-764-233A-1
: Sequence 1, Application US/08764233A
: Patent No. 5716849
:
: GENERAL INFORMATION:
:
: APPLICANT: Ligon, James M.
: APPLICANT: Schnupp, Thomas
: APPLICANT: Beck, James J.
: APPLICANT: Hill, Dwight S.
: APPLICANT: Neff, Snezana
: APPLICANT: Ryals, John A.
:
: TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
:
: NUMBER OF SEQUENCES: 10
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 520 White Plains Road, P.O. Box 2005
: CITY: Tarrytown
: STATE: NY
: COUNTRY: USA
: ZIP: 10591
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/764,233A
: FILING DATE:
:
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/729,214
: FILING DATE: 09-OCT-1996
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/258,261
: FILING DATE: 08-JUN-1994
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Meigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: 1506/CIP6
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (919) 541-8587
: TELEFAX: (919) 541-8689
:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 49377 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
:
: ORIGINAL SOURCE:
: ORGANISM: Sorangium cellulosum
: IMMEDIATE SOURCE:
: CLONE: p98/1, pUL3, and pVKM15
:
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 383..760

```

```

1  OTHER INFORMATION: /product= "SorB"
2  OTHER INFORMATION: /note= "This gene encodes a protein that is highly homolog
3  OTHER INFORMATION: the reductase domains of type I PKs such as eryA from
4  OTHER INFORMATION: Saccharopolyspora erythraea."
5  FEATURE:
6  NAME/KEY: misc_feature
7  LOCATION: 927..19874
8  OTHER INFORMATION: /product= "SorA"
9  OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs t
10 OTHER INFORMATION: are known to be involved in the synthesis of polypeptide
11 OTHER INFORMATION: compounds."
12 FEATURE:
13 NAME/KEY: misc_feature
14 LOCATION: 942..7115
15 OTHER INFORMATION: /product= "Module 1 of SorA"
16 FEATURE:
17 NAME/KEY: misc_feature
18 LOCATION: 7203..12884
19 OTHER INFORMATION: /product= "Module 2 of SorA"
20 FEATURE:
21 NAME/KEY: misc_feature
22 LOCATION: 13455..19615
23 OTHER INFORMATION: /product= "Module 3 of SorA"
24 FEATURE:
25 NAME/KEY: misc_feature
26 LOCATION: 19871..46318
27 OTHER INFORMATION: /product= "SorB"
28 OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs ge
29 FEATURE:
30 NAME/KEY: misc_feature
31 LOCATION: 19870..24556
32 OTHER INFORMATION: /product= "Module 1 of SorB"
33 FEATURE:
34 NAME/KEY: misc_feature
35 LOCATION: 24638..30820
36 OTHER INFORMATION: /product= "Module 2 of SorB"
37 FEATURE:
38 NAME/KEY: misc_feature
39 LOCATION: 30881..35446
40 OTHER INFORMATION: /product= "Module 3 of SorB"
41 FEATURE:
42 NAME/KEY: misc_feature
43 LOCATION: 35528..40114
44 OTHER INFORMATION: /product= "Module 4 of SorB"
45 FEATURE:
46 NAME/KEY: misc_feature
47 LOCATION: 40190..46318
48 OTHER INFORMATION: /product= "Module 5 of SorB"
49 FEATURE:
50 NAME/KEY: misc_feature
51 LOCATION: 46851..47891
52 OTHER INFORMATION: /product= "SorM"
53 OTHER INFORMATION: /note= "The protein encoded by the sorM gene is highly
54 OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
55 OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the
56 OTHER INFORMATION: polyketide rapamycin."
57 US-08-764-233A-1
58
59 Query Match 3.5% Score 43.6; DB 2: Length 49377;
60 Best Local Similarity 44.8% Pred. No. 0.11; Mismatches 354; Indels 4; Gaps 3;
61 Matches 290; Conservative 0;
62
63 Oy 148 GCCCATGAGTGCCTCATCCAGCAGAGTGCCGAGAGGGCTCGCTATCCAGCAGCGGAT 207
64 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
65 Db 41609 GCCCCAGCAGCAGCTACCCCTCGGATGTGCGCTATTGCTGTCGACACCCGCGCCOAC 41668
66 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
67 Oy 208 TTCAGGGGGGAGGGGCGCCCACTGTCGGCAGCAGCCCGCAGCAGAGTGTGGGCGGAGCTGAT 267
68 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
69 Db 41669 TTCGAGCAGCCGCGGCTCTCTGTAGCCCAACAGCGGAGCTCTCTCGCGCTCGAC 41728
70 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
71 Oy 268 TTATGTCTCAAGGTCAA-AGAACCGATAGCGGCGGAATAGCGCGCTCGAGACGGGCG 326

```

DB 41729 TCGGTGCGCCAGGACAAAGCCCGCCGAGACCGTCTCTGAGAGGAGCGGAGCAACGCGC 41788  
QY 327 ATCTGTTACAGCTTCTTGACATTTGGCCGCTCACGTCTTGACCGATGGCTGTGGAT 386  
DB 41789 AAGTCTCTCTTCTTCTTCTTGAGGAGGCTGCAAGTGGAGGAGGATGGCTCTGCTG 41848  
QY 387 TCGGAGCAGCAGTCAATTTGCTTACAGACCGCTCAGACCGCCGAGCGGAGCTACCTG 446  
DB 41849 CTGACACTCTCGCCGCTTCTCGGCTCAGCTGCAAGCAGGAGCGGCTGCTGCT 41908  
QY 447 CTGCGCCGATGAGCAGGATGCGCGGTGCACTGCGCCGCAAGTTGGCTTACAC--C 504  
DB 41909 CAGCTCAGTGAAGCCGCTGCGCTGCTGCGCGCAGCAGAGGCGCCCTCTCTGAC 41968  
QY 505 TGAATGAGACCCAAAGGGGGGCGGCTGTGATGAGGCGGAGGCTGCGGAGCGG 564  
DB 41969 CGCTGCACTGCTGACAGCCCGCTTTCGCTCATGCTCTCTGCGGCTCTG 42028  
QY 565 CCGAGC--TCGTGATGATGCGCGCGGACCGCGCTACAAAGCAGCCGCGATCGCCAC 623  
DB 42029 CGCTGCTGCGAGGAGAGCCGCGCGCTGCTGCGGCAAGTCAAGGAGATCGCGCC 42088  
QY 624 GCGATGGCGCGGAGCTTACGCTTACATCAACATCAAGCAACTTGGCAACTGAC 683  
DB 42089 GCTTCTGCGAGGCGCTCTCTCTGAGAGACCGCGCGCTACGCGCGCTGCGAC 42148  
QY 684 GCCGAGTCTGCGCGCGGATCCACACTGCTACTCATGCGCTTACGAGCTCGAGGCTG 743  
DB 42149 AAAGCGCTCACACCGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 42208  
QY 744 GTCAAGTGGCGAGCTGATGATGAGGCGGCTCTCTGCTGCGGAGGCGG 791  
DB 42209 GACCTCAGACCTACCGCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 42256

## RESULT 13

US-08-403-852D-6

Sequence 6, Application US/08403852D

Patent No. 5891695

GENERAL INFORMATION:

APPLICANT: Blanc, Veronique

APPLICANT: Blanc, Francis

APPLICANT: Crouzet, Joel

APPLICANT: Jacques, Nathalie

APPLICANT: Lacroix, Patricia

APPLICANT: Thibaut, Denis

APPLICANT: Zagorec, Monique

APPLICANT: Debussche, Laurent

APPLICANT: De Crey-Lagard, Valerie

APPLICANT: Polysynthesis of Streptogramins, Nucleotide Sequences

TITLE OF INVENTION: Coding for These Polypeptides And Their Use

NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:

ADDRESS: Finnegan, Henderson, Farbow, Garrett &amp; Dunner

STREET: 1300 I Street, N.W., Suite 700

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/403,852D

FILING DATE: 10-MAY-1995

PRIOR APPLICATION NUMBER: PCT/FR 93/00923

FILING DATE: 25-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 92/11441

FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806, 0054-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1833 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.pristinaespiralis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 103..1689  
US-08-403-852D-6

Query Match 3.4%; Score 42.4; DB 3; Length 1833;  
Best Local Similarity 50.5%; Pred. No. 0.09;  
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 532 TGCTGATGGCGGCGGCTGCGGCGGCTGCAACCGCGGCTGCGGATGCGGCGGCGCA 591  
DB 413 TGATGCTGCGAGCTGCTGCTTCTGCGCGCGGCTGCGGCGGCTGCGGCGGCGGAGCT 472  
QY 592 CCGCGGCTACAAAGCAGCGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGG 651  
DB 473 CCGCGGCTACAAAGCAGCGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGG 532  
QY 652 ACATCAACATGACAAACTTGGCAACTGAGCGGCGGCTGCGGCGGCTGCGGCGGCTG 711  
DB 533 CCAACCTGCGGCTGCACTGAGGCTTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGC 592  
QY 712 GCTACTATGCGGCTTACGAGCTCG 735  
DB 593 TGTGGGAGCGGCGGCTTACCTCG 515

## RESULT 14

US-08-173-508-3

Sequence 3, Application US/08173508

Patent No. 5616485

GENERAL INFORMATION:

APPLICANT: Bartfield, Daniel

APPLICANT: Butler, Michael J.

APPLICANT: Hadary, Dany

APPLICANT: Jenish, David

APPLICANT: Krieger, Timothy

APPLICANT: Streptomyces proteases and improved

TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED

TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND

TITLE OF INVENTION: POLYPEPTIDES

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESS: Foley &amp; Lardner

STREET: 3000 K Street, N.W.

CITY: Washington, D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/173,508



FILED DATE: 23-DEC-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 18740/125/CACO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202 672 5300  
TELEFAX: 202 672 5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2185 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 531..2069  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 531..902  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 903..2069  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 531..533  
OTHER INFORMATION: /note="Met at position -124  
US-08-173-508-3  
represents fMet"

Query Match 3.4%; Score 42.4; DB 1; Length 2185;  
Best Local Similarity 47.4%; Pred. No. 0.094;  
Matches 127; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 524 CCGCGGTGCTGATGCGGGGGTCCCGGCGTCAACCGCGCGATGCTGTGATCGG 583  
DB 1250 CCGCGGTGCTGATGCGGGGGTCCCGGCGTCAACCGCGCGATGCTGTGATCGG 1309  
QY 584 CCGCGGACCGCGCGGCTTACACGACGCGCGATGCGCGAAGCGCGCGATGCGG 643  
DB 1310 CCGCGGACCGCGGCTTACACGACGCGCGATGCGCGAAGCGCGCGATGCGG 1369  
QY 644 GGTTCAGACATCAACATGACAACTTCGCGCACTCGACGCGCGATGCGGCGGAT 703  
DB 1370 CCGCGGACCGCGGCTTACACGACGCGCGATGCGCGAAGCGCGCGATGCGG 1429  
QY 704 CCACACTGCTACTGATGCGGCGTACGAGGCTGCGCGAAGCGCGATGCGG 763  
DB 1430 CCACACTGCTACTGATGCGGCGTACGAGGCTGCGCGAAGCGCGATGCGG 1489  
QY 764 GATTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 791  
DB 1490 GAAAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1517

RESULT 15  
US-08-265-310-3  
Sequence 3, Application US/08265310  
Patent No. 5856166

GENERAL INFORMATION:  
APPLICANT: Bartfeld, Daniel J.  
APPLICANT: Butler, Michael J.  
APPLICANT: Hadary, Dany  
APPLICANT: Jenish, David  
APPLICANT: Krieger, Timothy  
APPLICANT: Malek, Lawrence T.  
APPLICANT: Soostmeyer, Gisela  
APPLICANT: Walczyk, Eva  
APPLICANT: Krysman, Phyllis  
APPLICANT: Garven, Sheila

TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED  
TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND  
TITLE OF INVENTION: POLYPEPTIDES  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/265,310  
FILING DATE: 24-JUN-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,508  
FILING DATE: 23-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 18740/133/CACO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202 672 5300  
TELEFAX: 202 672 5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2185 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 531..2069  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 531..902  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 903..2069  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 531..533  
OTHER INFORMATION: /note="Met at position -124  
US-08-265-310-3  
represents fMet"

Query Match 3.4%; Score 42.4; DB 3; Length 2185;  
Best Local Similarity 47.4%; Pred. No. 0.094;  
Matches 127; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 524 CCGCGGTGCTGATGCGGGGGTCCCGGCGTCAACCGCGCGATGCTGTGATCGG 583  
DB 1250 CCGCGGTGCTGATGCGGGGGTCCCGGCGTCAACCGCGCGATGCTGTGATCGG 1309  
QY 584 CCGCGGACCGCGCGGCTTACACGACGCGCGATGCGCGAAGCGCGCGATGCGG 643  
DB 1310 CCGCGGACCGCGGCTTACACGACGCGCGATGCGCGAAGCGCGCGATGCGG 1369  
QY 644 GGTTCAGACATCAACATGACAACTTCGCGCACTCGACGCGCGATGCGGCGGAT 703  
DB 1370 CCGCGGACCGCGGCTTACACGACGCGCGATGCGCGAAGCGCGCGATGCGG 1429  
QY 704 CCACACTGCTACTGATGCGGCGTACGAGGCTGCGCGAAGCGCGATGCGG 763  
DB 1430 CCACACTGCTACTGATGCGGCGTACGAGGCTGCGCGAAGCGCGATGCGG 1489

Fri Jun 23 09:31:34 2000

us-09-362-485-6.rni

Page 12

QY 764 GATTGGGGCCGCTCTGTGCCAGCGCC 791  
||| ||| ||| ||| ||| |||  
Db 1490 GAAGGGCGCGGTACACGTCCGCGCGGCC 1517

Search completed: June 22, 2000, 15:11:10  
Job time: 18148 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compygen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:07:36 ; Search time 5541.94 Seconds  
(without alignments)  
903.245 Million cell updates/sec

Title: US-09-362-485-6  
Perfect score: 1235  
Sequence: 1 ACTTGACATATATCGAAC.....GGCAGCGATGATCGGCC 1235

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: gb\_est33:\*  
53: gb\_est34:\*  
54: gb\_est35:\*  
55: gb\_est36:\*  
56: gb\_est37:\*  
57: gb\_est38:\*  
58: gb\_est39:\*  
59: gb\_est40:\*  
60: gb\_est41:\*  
61: gb\_est42:\*  
62: gb\_est43:\*  
63: gb\_est44:\*  
64: gb\_est45:\*  
65: gb\_est46:\*  
66: gb\_est47:\*  
67: gb\_est48:\*  
68: gb\_est49:\*  
69: gb\_est50:\*  
70: gb\_est51:\*  
71: gb\_est52:\*  
72: gb\_est53:\*  
73: gb\_est54:\*  
74: gb\_est55:\*  
75: gb\_est56:\*  
76: gb\_est57:\*  
77: gb\_est58:\*  
78: gb\_est59:\*  
79: gb\_est60:\*  
80: gb\_est61:\*  
81: gb\_est62:\*  
82: gb\_est63:\*  
83: gb\_est64:\*  
84: gb\_est65:\*  
85: gb\_est66:\*  
86: gb\_est67:\*  
87: gb\_est68:\*  
88: gb\_est69:\*  
89: gb\_est70:\*  
90: gb\_est71:\*  
91: gb\_est72:\*  
92: gb\_est73:\*  
93: gb\_est74:\*  
94: gb\_est75:\*  
95: gb\_est76:\*  
96: gb\_est77:\*  
97: gb\_est78:\*  
98: gb\_est79:\*  
99: gb\_est80:\*  
100: gb\_est81:\*  
101: gb\_est82:\*  
102: gb\_est83:\*  
103: gb\_est84:\*  
104: gb\_est85:\*  
105: gb\_est86:\*  
106: gb\_est87:\*  
107: gb\_est88:\*  
108: gb\_est89:\*  
109: gb\_est90:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result

Query

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compyen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 18:04:14 ; Search time 123.79 Seconds

(Without alignments)  
15.751 Million cell updates/sec

Title: US-09-362-485-28

Sequence: 1 AACGAATTCGGGTG 15

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

Issued Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/1na/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/1na/5C\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/1na/5D\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/1na/6\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/1na/PTCUS\_COMB.seq: \*  
7: /cgn2\_6/ptodata/2/1na/Backfillseq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13.4	89.3	206	US-08-644-034A-5	Sequence 5, Appli
2	13	86.7	80	US-08-468-275-2	Sequence 2, Appli
3	13	86.7	207	US-08-142-551B-8	Sequence 8, Appli
4	12.4	82.7	48	US-09-211-631-20	Sequence 20, Appli
5	12.4	82.7	153	US-08-651-136C-49	Sequence 49, Appli
6	12.4	82.7	1389	US-08-702-344-20	Sequence 20, Appli
7	12.4	82.7	4145	US-08-314-917-1	Sequence 1, Appli
8	12.4	82.7	4145	US-08-265-046-1	Sequence 1, Appli
9	12.4	82.7	4145	US-08-465-522-1	Sequence 1, Appli
10	12.4	82.7	4145	PCT-US93-11401-1	Sequence 1, Appli
11	12.4	82.7	4145	PCT-US95-07849-1	Sequence 1, Appli
12	12.4	82.7	35	US-08-173-510B-51	Sequence 51, Appli
13	12	80.0	35	US-08-448-218-51	Sequence 51, Appli
14	12	80.0	35	US-08-450-457-51	Sequence 51, Appli
15	12	80.0	35	US-08-540-406-1	Sequence 1, Appli
16	12	80.0	736	US-08-656-055-1	Sequence 1, Appli
17	12	80.0	736	PCT-US95-13233-1	Sequence 1, Appli
18	12	80.0	1699	US-08-484-933B-3	Sequence 3, Appli
19	12	80.0	1699	US-08-484-158B-3	Sequence 3, Appli
20	12	80.0	1699	US-08-484-550A-3	Sequence 3, Appli
21	12	80.0	1699	US-08-480-150A-3	Sequence 3, Appli
22	12	80.0	1699	US-08-458-731-3	Sequence 3, Appli
23	12	80.0	1699	US-08-149-223A-3	Sequence 3, Appli
24	12	80.0	5864	US-08-894-440-4	Sequence 4, Appli
25	11.8	78.7	171	US-08-644-034A-4	Sequence 4, Appli
26	11.8	78.7	224	US-08-248-474-48	Sequence 48, Appli
27	11.8	78.7	326	US-08-111-316-3	Sequence 3, Appli

28	11.8	78.7	326	1	US-08-111-316-5	Sequence 5, Appli
29	11.8	78.7	326	1	US-08-468-405-3	Sequence 3, Appli
30	11.8	78.7	326	1	US-08-468-405-5	Sequence 5, Appli
31	11.8	78.7	348	3	US-08-642-229A-10	Sequence 10, Appli
32	11.8	78.7	375	1	US-08-111-316-4	Sequence 4, Appli
33	11.8	78.7	375	1	US-08-468-405-4	Sequence 4, Appli
34	11.8	78.7	429	5	US-08-520-550A-47	Sequence 47, Appli
35	11.8	78.7	476	1	US-08-111-316-2	Sequence 2, Appli
36	11.8	78.7	476	1	US-08-468-405-2	Sequence 2, Appli
37	11.8	78.7	560	5	US-08-520-550A-46	Sequence 46, Appli
38	11.8	78.7	580	3	US-08-485-778-45	Sequence 45, Appli
39	11.8	78.7	580	5	US-08-520-550A-43	Sequence 43, Appli
40	11.8	78.7	605	3	US-08-769-967A-33	Sequence 33, Appli
41	11.8	78.7	1128	2	US-08-815-688A-4	Sequence 4, Appli
42	11.8	78.7	1246	1	US-08-446-777-3	Sequence 3, Appli
43	11.8	78.7	1312	1	US-08-205-506A-1	Sequence 1, Appli
44	11.8	78.7	1312	6	PCT-US94-02389-1	Sequence 1, Appli
45	11.8	78.7	1313	3	US-08-463-911-6	Sequence 6, Appli

#### ALIGNMENTS

RESULT 1  
US-08-644-034A-5  
Sequence 5, Application US/08644034A  
Patent No. 5882862  
GENERAL INFORMATION:  
APPLICANT: Howard B. Lieberman and Kevin M. Hopkins  
TITLE OF INVENTION: A HUMAN RADIORESISTANCE/CELL CYCLE  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham, LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/644,034A  
FILING DATE: May 9, 1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: John P. White  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0575/50307  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 206 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-08-644-034A-5

Query Match 89.3% Score 13.4; DB 3; Length 206;  
Best Local Similarity 93.3% Pred. NO. 29;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AACGAATTCGGGTG 15  
DB 44 AAGCATTCGGGTG 58

RESULT 2  
US-08-468-275-2  
Sequence 2, Application US/08468275  
Patent No. 5747453  
GENERAL INFORMATION:  
APPLICANT: HOLADAY, LESLIE A.  
APPLICANT: OLDENBURG, KEVIN R.  
TITLE OF INVENTION: METHOD FOR INCREASING THE  
NUMBER OF INVENTION: ELECTROTRANSPORT FLOW OF POLYPEPTIDES  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ALZA CORPORATION  
STREET: 950 PAGE MILL ROAD  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94303-0802  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,275  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: MILLER, D. BYRON  
REGISTRATION NUMBER: 30,661  
REFERENCE/DOCKET NUMBER: 0360-0002; ARC-2349  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 496-8150  
TELEFAX: (415) 496-8048  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-468-275-2

Query Match 86.7%; Score 13; DB 2; Length 80;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACGAATCCGGGT 14  
|||||  
Db 59 ACGAATCCGGGT 71

RESULT 3  
US-08-142-551B-8/c  
Sequence 8, Application US/08142551B  
Patent No. 5814603  
GENERAL INFORMATION:  
APPLICANT: Oldenburg, Kevin R.  
APPLICANT: Selick, Harold E.  
TITLE OF INVENTION: COMPOUNDS WITH PTH ACTIVITY AND  
NUMBER OF INVENTION: RECOMBINANT DNA VECTORS ENCODING SAME  
NUMBER OF SEQUENCES: 132  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: 699 Prince Street  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: US  
ZIP: 22313  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/142,551B  
FILING DATE: 25-OCT-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/077,296  
FILING DATE: 14-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/898,219  
FILING DATE: 12-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/965,677  
FILING DATE: 22-OCT-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Swiss, Gerald F.  
REGISTRATION NUMBER: 30,113  
REFERENCE/DOCKET NUMBER: 000324-010  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 854-7400  
TELEFAX: (415) 854-8275  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 207 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..201  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..51  
OTHER INFORMATION: /note="Encodes the leader peptide  
OTHER INFORMATION: sequence that serves to direct the protein into  
OTHER INFORMATION: inclusion bodies."  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 70..174  
OTHER INFORMATION: /note="Encodes the protein or  
OTHER INFORMATION: peptide of interest"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 175..207  
OTHER INFORMATION: /note="Encodes amino acid sequence  
OTHER INFORMATION: having six histamines that serves as a tag for the  
OTHER INFORMATION: purification of the protein on a nickel column."  
US-08-142-551B-8

Query Match 86.7%; Score 13; DB 2; Length 207;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACGAATCCGGGT 14  
|||||  
Db 53 ACGAATCCGGGT 41

RESULT 4  
US-09-211-631-20/c  
Sequence 20, Application US/09211631  
Patent No. 6001597  
GENERAL INFORMATION:  
APPLICANT: Raymond, Christopher K.  
APPLICANT: Vanaja, Erica  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PRODUCING  
HETEROLOGOUS POLYPEPTIDES IN PICHIA METHANOLICA  
FILE REFERENCE: 96-16C2  
CURRENT APPLICATION NUMBER: US/09/211,631  
CURRENT FILING DATE: 1998-12-15

NUMBER OF SEQ ID NOS: 23  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 20  
LENGTH: 48  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: oligonucleotide primer  
US-09-211-631-20

Query Match 82.7%; Score 12.4; DB 5; Length 48;  
Best Local Similarity 92.9%; Pred. No. 1e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACGAATTCGGGT 14  
DB 21 ACGAATTCGGGT 8

## RESULT 5

US-08-651-136C-49  
Sequence 49, Application US/08651136C  
Patent No. 6001639

GENERAL INFORMATION:  
APPLICANT: Schuelein, Martin  
APPLICANT: Andersen, Lene N.  
APPLICANT: Lassen, Soren F.  
APPLICANT: Kauppinen, Markus S.  
APPLICANT: Lange, Lene  
APPLICANT: Nielsen, Ruby I.  
APPLICANT: Ihara, Michiko  
APPLICANT: Takagi, Shinobu  
TITLE OF INVENTION: No. 6001639e1 Endoglucanases  
NUMBER OF SEQUENCES: 109  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6001639e No. 6001639e1sk of No. 6001639th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/651,136C  
FILING DATE: 21-MAY-1996  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4366.200-US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 49:

SEQUENCE CHARACTERISTICS:

LENGTH: 153 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..153

US-08-651-136C-49

Query Match 82.7%; Score 12.4; DB 5; Length 153;  
Best Local Similarity 92.9%; Pred. No. 1.1e+02;

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACGAATTCGGGTG 15  
DB 37 ACGAATTCGGGTG 50

## RESULT 6

US-08-702-344-20  
Sequence 20, Application US/08702344  
Patent No. 572315

GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John  
APPLICANT: Lavallee, Edward  
APPLICANT: Racine, Lisa  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Spaulding, Vikki  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/702,344

FILING DATE:

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A.

REGISTRATION NUMBER: 32,724

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8224

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 1389 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-702-344-20

Query Match 82.7%; Score 12.4; DB 2; Length 1389;  
Best Local Similarity 92.9%; Pred. No. 1.1e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACGAATTCGGGTG 15  
DB 431 ACGAATTCGGGTG 444

RESULT 7  
US-08-314-917-1  
Sequence 1, Application US/08314917  
Patent No. 546830

GENERAL INFORMATION:  
APPLICANT: Billiar, Timothy R.  
APPLICANT: Nussler, Andreas K.  
APPLICANT: Geller, David A.  
APPLICANT: Simmons, Richard L.  
TITLE OF INVENTION: cDNA Clone for Human Inducible Nitric

TITLE OF INVENTION: Oxide Synthase And Process for Preparing Same  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold B. Silverman  
ADDRESSEE: Eckert Seamans Cherin & Mellott  
STREET: 600 Grant Street, 42nd Floor  
CITY: Pittsburgh  
STATE: PA  
COUNTRY: USA  
ZIP: 15219  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/314,917  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/981,344  
FILING DATE: 25-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Silverman, Arnold B.  
REGISTRATION NUMBER: 22,614  
REFERENCE/DOCKET NUMBER: 116972  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (412) 566-6000  
TELEFAX: (412) 566-6099  
TELEX: 866172  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4145 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
DESCRIPTION: Human Hepatocyte Inducible Nitric Oxide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
TISSUE TYPE: Induced Human Hepatocyte RNA  
IMMEDIATE SOURCE:  
LIBRARY: Lambda Zap II cDNA  
CLONE: PHINOS  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: unknown  
MAP POSITION: unknown  
UNITS: unknown  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 207..3668  
IDENTIFICATION METHOD: Experiment  
US-08-314-917-1

Query Match 82.7%; Score 12.4; DB 1; Length 4145;  
Best Local Similarity 92.9%; Pred. No. 1.1e+2;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACGAATTCGGGGTG 15  
|||||

DB 970 ACGACTTCGGGGTG 983

RESULT 8  
US-08-265-046-1  
; Sequence 1, Application US/08265046  
; Patent No. 5658565  
; GENERAL INFORMATION:  
; APPLICANT: Timothy R. Billiar  
; APPLICANT: Edith Tzeng

APPLICANT: Andreas K. Nussler  
APPLICANT: David A. Geller  
APPLICANT: Richard L. Simmons  
TITLE OF INVENTION: Inducible Nitric Oxide Synthase  
TITLE OF INVENTION: Gene for Treatment of Disease  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lewis F. Gould, Jr.  
ADDRESSEE: Eckert Seamans Cherin & Mellott  
STREET: 1700 Market Street, Suite 3232  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/265,046  
FILING DATE: 24-JUN-1994  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Gould, Lewis F. Jr.  
REGISTRATION NUMBER: 25,057  
REFERENCE/DOCKET NUMBER: 119130  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 575-6020  
TELEFAX: (215) 575-6015  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4145 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
DESCRIPTION: Human Hepatocyte Inducible Nitric Oxide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
TISSUE TYPE: Induced Human Hepatocyte RNA  
IMMEDIATE SOURCE:  
LIBRARY: Lambda Zap II cDNA  
CLONE: PHINOS  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: unknown  
MAP POSITION: unknown  
UNITS: unknown  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 207..3668  
IDENTIFICATION METHOD: Experiment  
US-08-265-046-1

Query Match 82.7%; Score 12.4; DB 1; Length 4145;  
Best Local Similarity 92.9%; Pred. No. 1.1e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACGAATTCGGGGTG 15  
|||||

DB 970 ACGACTTCGGGGTG 983

RESULT 9  
US-08-465-522-1  
; Sequence 1, Application US/08465522  
; Patent No. 5882908  
; GENERAL INFORMATION:  
; APPLICANT: Billiar, Timothy R.



APPLICANT: Nusaler, Andreas K.  
APPLICANT: Geller, David A.  
APPLICANT: Simmons, Richard L.  
TITLE OF INVENTION: CDNA Clone for Human Inducible Nitric  
Oxide Synthase And Process for Preparing Same  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lewis F. Gould, Jr.  
ADDRESSEE: Eckert Seamans Cherin & Mellott  
STREET: 1700 Market St. Suite 3232  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,522  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gould, Jr., Lewis F.  
REGISTRATION NUMBER: 25,057  
REFERENCE/DOCKET NUMBER: 116972-6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 575-6020  
TELEFAX: (215) 575-6015  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4145 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
DESCRIPTION: Human Hepatocyte Inducible Nitric Oxide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
TISSUE TYPE: Induced Human Hepatocyte RNA  
IMMEDIATE SOURCE:  
LIBRARY: Lambda Zap II CDNA  
CLONE: PHINOS  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: unknown  
MAP POSITION: unknown  
UNITS: unknown  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 207..3668  
IDENTIFICATION METHOD: Experiment  
US-08-465-522-1

Query Match 82.7%; Score 12.4; DB 3; Length 4145;  
Best Local Similarity 92.9%; Pred. No. 1.1e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACGAATTCGGGTG 15  
DB 970 ACGACTTCGGGTG 983

RESULT 10  
PCT-US93-11401-1  
Sequence 1, Application PCTUS9311401  
GENERAL INFORMATION:  
APPLICANT: Billiar, Timothy R.  
APPLICANT: Nusaler, Andreas K.

APPLICANT: Geller, David A.  
APPLICANT: Simmons, Richard L.  
TITLE OF INVENTION: CDNA Clone for Human Inducible Nitric  
Oxide Synthase And Process for Preparing Same  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold B. Silverman  
ADDRESSEE: Eckert Seamans Cherin & Mellott  
STREET: 600 Grant Street, 42nd Floor  
CITY: Pittsburgh  
STATE: PA  
COUNTRY: USA  
ZIP: 15219  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/11401  
FILING DATE: 25-NOV-1992  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/981,344  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Silverman, Arnold B.  
REGISTRATION NUMBER: 22,614  
REFERENCE/DOCKET NUMBER: 116972  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (412) 566-6000  
TELEFAX: (412) 566-6099  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4145 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
TISSUE TYPE: Induced Human Hepatocyte RNA  
IMMEDIATE SOURCE:  
LIBRARY: Lambda Zap II CDNA  
CLONE: PHINOS  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: unknown  
MAP POSITION: unknown  
UNITS: unknown  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 207..3668  
IDENTIFICATION METHOD: Experiment  
PCT-US93-11401-1

Query Match 82.7%; Score 12.4; DB 6; Length 4145;  
Best Local Similarity 92.9%; Pred. No. 1.1e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACGAATTCGGGTG 15  
DB 970 ACGACTTCGGGTG 983

RESULT 11  
PCT-US95-07849-1  
Sequence 1, Application PC/TUS9507849  
GENERAL INFORMATION:  
APPLICANT: University of Pittsburgh of the Commonwealth System of Higher  
APPLICANT: Education

TITLE OF INVENTION: Inducible Nitric Oxide Synthase  
TITLE OF INVENTION: Gene for Treatment of Disease  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lewis F. Gould, Jr.  
ADDRESSEE: Eckert Seamans Cherin & Mellott  
STREET: 1700 Market Street, Suite 3232  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/07849  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gould, Lewis F. Jr.  
REGISTRATION NUMBER: 25,057  
REFERENCE/DOCKET NUMBER: 119130-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 575-6020  
TELEFAX: (215) 575-6015  
TELEX:  
INFORMATION FOR SEQ. ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4145 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
DESCRIPTION: Human Hepatocyte Inducible Nitric Oxide  
HYPOTHEtical: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
TISSUE TYPE: Induced Human Hepatocyte RNA  
IMMEDIATE SOURCE:  
LIBRARY: Lambda Zap II cDNA  
CLONE: PHINOS  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: unknown  
MAP POSITION: unknown  
UNITS: unknown  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 207..3668  
IDENTIFICATION METHOD: Experiment  
PCT-US95-07849-1

Query Match 82.7%; Score 12.4; DB 6; Length 4145;  
Best Local Similarity 92.9%; Pred. No. 1.1e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACGAATCCGGG 15  
DB 970 ACGACTTCGGG 983

RESULT 12  
US-08-173-510B-51/c  
Sequence 51, Application US/08173510B  
Patent No. 5747296  
GENERAL INFORMATION:  
APPLICANT: MATTHEW MOYLE, ET AL.  
TITLE OF INVENTION: NOVEL NEUTROPHIL INHIBITORS  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: Storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/173,510B  
FILING DATE: 23-DEC-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/151,064  
FILING DATE: 10-NOV-1993  
APPLICATION NUMBER: 08/060,433  
FILING DATE: 11-MAY-1993  
APPLICATION NUMBER: 07/996,972  
FILING DATE: 24-DEC-1992  
APPLICATION NUMBER: 07/881,721  
FILING DATE: 11-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 205/073  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ. ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: NUCLEIC  
US-08-173-510B-51

Query Match 80.0%; Score 12; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACGAATCCGG 12  
DB 12 ACGAATCCGG 1

RESULT 13  
US-08-458-218-51/c  
Sequence 51, Application US/08458218  
Patent No. 5789178  
GENERAL INFORMATION:  
APPLICANT: MATTHEW MOYLE ET AL.  
TITLE OF INVENTION: NOVEL NEUTROPHIL INHIBITORS  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: Storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0

SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,218  
FILING DATE: 11-MAY-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/151,064  
FILING DATE: 10-NOVEMBER-1993  
APPLICATION NUMBER: 08/060,433  
FILING DATE: 11-MAY-1993  
APPLICATION NUMBER: 07/881,721  
FILING DATE: 11-MAY-1992  
APPLICATION NUMBER: 07/996,972  
FILING DATE: 24-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 203/726  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: NUCLEIC  
US-08-458-218-51

Query Match 80.0%; Score 12; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGAATTCGG 12  
|||||  
DB 12 AACGAATTCGG 1

RESULT 14  
US-08-450-497-51/C  
Sequence 51, Application US/08450497  
Patent No. 5919900  
GENERAL INFORMATION:  
APPLICANT: MATTHEW MOYLE, ET AL.  
TITLE OF INVENTION: NOVEL NEUTROPHIL INHIBITORS  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,497  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/173,510  
FILING DATE: 23-DEC-1993  
APPLICATION NUMBER: 08/151,064  
FILING DATE: 10-NOV-1993  
APPLICATION NUMBER: 08/060,433  
FILING DATE: 11-MAY-1993

APPLICATION NUMBER: 07/996,972  
FILING DATE: 24-DEC-1992  
APPLICATION NUMBER: 07/881,721  
FILING DATE: 11-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 205/073  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: NUCLEIC  
US-08-450-497-51

Query Match 80.0%; Score 12; DB 4; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGAATTCGG 12  
|||||  
DB 12 AACGAATTCGG 1

RESULT 15  
US-08-540-406-1/C  
Sequence 1, Application US/08540406  
Patent No. 5337538  
GENERAL INFORMATION:  
APPLICANT: SCOTT, MATHEW P  
APPLICANT: GOODRICH, LISA V  
APPLICANT: JOHNSON, RONALD L  
TITLE OF INVENTION: Patched Genes and their use  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/540,406  
FILING DATE: 06-OCT-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rowland, Bertram I  
REGISTRATION NUMBER: 20015  
REFERENCE/DOCKET NUMBER: a60190-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 736 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-540-406-1

Query Match 80.0%; Score 12; DB 3; Length 736;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0;

QY 4 GAATTCGGGTG 15  
 |||||  
 Db 147 GAATTCGGGTG 136

Search completed: June 22, 2000, 18:04:15  
 Job time: 9722 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:50:07 ; Search time 8627.09 Seconds

(Without alignments)  
-140.386 Million cell updates/sec

Title: US-09-362-485-2  
Perfect score: 1245  
Sequence: 1 ATCTGCAGATATATCGAAC.....AAGCGATGATGCGCGCCG 1245

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenBankl:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_om:\*  
5: gb\_om:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pl4:\*  
11: gb\_pl5:\*  
12: gb\_pl6:\*  
13: gb\_pl7:\*  
14: gb\_pl8:\*  
15: gb\_pl9:\*  
16: gb\_pl10:\*  
17: gb\_pl11:\*  
18: gb\_pl12:\*  
19: gb\_pl13:\*  
20: gb\_pl14:\*  
21: gb\_pl15:\*  
22: gb\_pl16:\*  
23: gb\_pl17:\*  
24: gb\_pl18:\*  
25: gb\_pl19:\*  
26: gb\_pl20:\*  
27: gb\_pl21:\*  
28: gb\_pl22:\*  
29: gb\_pl23:\*  
30: gb\_pl24:\*  
31: gb\_pl25:\*  
32: gb\_pl26:\*  
33: gb\_pl27:\*  
34: gb\_pl28:\*  
35: gb\_pl29:\*  
36: gb\_pl30:\*  
37: gb\_pl31:\*  
38: gb\_pl32:\*  
39: gb\_pl33:\*  
40: gb\_pl34:\*  
41: gb\_pl35:\*  
42: gb\_pl36:\*  
43: gb\_pl37:\*  
44: gb\_pl38:\*  
45: gb\_pl39:\*

45: gb\_hg9:\*  
46: em\_hg1:\*  
47: em\_hg2:\*  
48: em\_hg3:\*  
49: em\_hg4:\*  
50: gb\_pl3:\*  
51: gb\_pl4:\*  
52: gb\_hg8:\*  
53: gb\_hg9:\*  
54: gb\_hg10:\*  
55: gb\_hg11:\*  
56: gb\_hg12:\*  
57: gb\_hg13:\*  
58: gb\_hg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1245	100.0	1245	5 A87605	A87605 Sequence 2
2	1245	100.0	1245	5 A89745	A89745 Sequence 2
3	1245	100.0	1260	5 A87604	A87604 Sequence 1
4	1245	100.0	1260	5 A89744	A89744 Sequence 1
5	1245	100.0	2584	1 MTALADH	X63069 M. tuberculo
6	1221.4	98.1	56414	1 MTALADH	AL008967 Mycobacte
7	1220	98.0	1235	5 A87613	A87613 Sequence 10
8	1220	98.0	1235	5 A89753	A89753 Sequence 10
9	1219.4	97.9	1237	5 A87607	A87607 Sequence 4
10	1219.4	97.9	1237	5 A89747	A89747 Sequence 4
11	1216	97.7	1235	5 A87606	A87606 Sequence 3
12	1216	97.7	1235	5 A89746	A89746 Sequence 3
13	1210.4	97.2	1228	5 A87608	A87608 Sequence 5
14	1210.4	97.2	1228	5 A89748	A89748 Sequence 5
15	1207	96.9	1235	5 A87609	A87609 Sequence 6
16	1207	96.9	1235	5 A89749	A89749 Sequence 6
17	1206.4	96.9	1235	5 A87611	A87611 Sequence 8
18	1206.4	96.9	1235	5 A89751	A89751 Sequence 8
19	1199.4	96.3	1229	5 A87610	A87610 Sequence 7
20	1199.4	96.3	1229	5 A89750	A89750 Sequence 7
21	1192	95.7	1209	5 A87612	A87612 Sequence 9
22	1185	95.2	1208	5 A89752	A89752 Sequence 9
23	1178	94.6	1194	2 MTU92472	U92472 Mycobacteri
24	720	57.8	720	5 A89759	A89759 Sequence 16
25	522.4	42.0	40745	1 SC151	AL109848 Streptomy
26	467.6	37.6	682	5 A89758	A89758 Sequence 15
27	467.6	37.6	682	5 A89762	A89762 Sequence 19
28	361	29.0	8046	2 AF049107	AF049107 Mycobactu
29	337.2	27.1	11514	2 AE002028	AE002028 Deinococc
30	315.4	25.3	2780	1 RLE238118	AJ238118 Rhizobium
31	314.8	25.3	1585	2 AF070716	AF070716 Vibrio pr
32	296.8	23.8	1736	1 BACALDH	M33299 B. stearothe
33	277	22.2	1477	2 AF070715	AF070715 Shewanell
34	272.4	21.9	213	1 D37807	D37807 Phormidium
35	261.2	21.0	1125	5 E01847	E01847 DNA encodin
36	246.8	19.8	13740	1 D90900	D90900 Synecocyst
37	227.2	18.2	1375	1 AB013821	AB013821 Enterobac
38	227	18.2	1776	1 BACALDH	M33298 B. sphaericu
39	219.6	17.6	1709	1 BACALDH	L20916 Bacillus su
40	219.6	17.6	12316	1 BS282015	BS282015 B. subtilis
41	219.6	17.6	217420	1 BSUB0017	Z99120 Bacillus su
42	202.2	16.2	1714	2 AF070714	AF070714 Carnobact
43	108.8	8.7	10652	2 AE001565	AE001565 Helicobac
44	100.8	8.1	10512	2 AE000639	AE000639 Helicobac
45	66.4	5.3	3160	1 RR05294	U05294 Rhodospirill

## ALIGNMENTS

```

RESULT 1
LOCUS A87605 1245 bp DNA
DEFINITION Sequence 2 from Patent WO9836089.
ACCESSION A87605
VERSION A87605.1 GI:6736245
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1245)
AUTHORS Flohe, L. and Singh, M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A-20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
Source 1..1245
/db_xref="taxon:32644"
BASE COUNT 238 a 398 c 387 g 222 t
ORIGIN
Query Match 100.0%; Score 1245; DB 5; Length 1245;
Best Local Similarity 100.0%; Pred. No. 1.2e-171;
Matches 1245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTTGAGATTATCGAACTTCTTCACACTGAGAGGTACAGTATGAGAGTGTATATC 60
DB 1 ATCTTGAGATTATCGAACTTCTTCACACTGAGAGGTACAGTATGAGAGTGTATATC 60
QY 61 ATGGCGGTGCTATTCGACGACGACCAAAAAACAAGCAATTCATTCGGGTGGCCATC 120
DB 61 ATGGCGGTGCTATTCGACGACGACCAAAAAACAAGCAATTCATTCGGGTGGCCATC 120
QY 121 ACCCGGCGGCGTGGCGGAACTAACCGGTGTCGTCATGAGTGTATTCAGGCAAGT 180
DB 121 ACCCGGCGGCGTGGCGGAACTAACCGGTGTCGTCATGAGTGTATTCAGGCAAGT 180
QY 181 GCGGAGAGGGGCTGGGCTATCACGAGCGGATTCAGAGGGGAGGGGCGCACTGGTC 240
DB 181 GCGGAGAGGGGCTGGGCTATCACGAGCGGATTCAGAGGGGAGGGGCGCACTGGTC 240
QY 241 GGCACCGCGGACGAGGTGGGCGGACGCTGATTTATGTCACAGTCTTGGCAATGGCCATA 300
DB 241 GGCACCGCGGACGAGGTGGGCGGACGCTGATTTATGTCACAGTCTTGGCAATGGCCATA 300
QY 301 GCGGCGGAATACGGCGGCTGCGACGAGGAGATCTTGTTCACGTTCTTGCAATTTGGCC 360
DB 301 GCGGCGGAATACGGCGGCTGCGACGAGGAGATCTTGTTCACGTTCTTGCAATTTGGCC 360
QY 361 GCGTCAGTGTGTCACGATGCTGTTGATTCGCGACACGATGCTCAATTTGGCCATAGAG 420
DB 361 GCGTCAGTGTGTCACGATGCTGTTGATTCGCGACACGATGCTCAATTTGGCCATAGAG 420
QY 421 ACCGTCCAGACGCGCGAGGCGGCACTACCCCTGCTTGCCTCCCGATGAGCAAGTGCCTG 480
DB 421 ACCGTCCAGACGCGCGAGGCGGCACTACCCCTGCTTGCCTCCCGATGAGCAAGTGCCTG 480
QY 481 CGACTCGCGCGGCGGCTTACCACTGATGCGAACCCAAAGGGGCGCGGTGTG 540
DB 481 CGACTCGCGCGGCGGCTTACCACTGATGCGAACCCAAAGGGGCGCGGTGTG 540
QY 541 CTGATGGGCGGCGGTGCGCGGCTGCAACCGGACGCTGCTGATGCGGCGCGGACCG 600
DB 541 CTGATGGGCGGCGGTGCGCGGCTGCAACCGGACGCTGCTGATGCGGCGCGGACCG 600
QY 601 GCGGCGTACAAAGCGAGCGCGGATGCCAAGCGCATGGGCGGACGCTTACGTTAGAC 660
DB 601 GCGGCGTACAAAGCGAGCGCGGATGCCAAGCGCATGGGCGGACGCTTACGTTAGAC 660
QY 661 ATCAACATCGACAAACTTGGCACTGAGCGCCAGATTTCGGCGCGGATTCACACTGCC 720

```

```

DB 661 ATCAACATCGACAAACTTGGCACTGAGCGCCGAGTTCTGCGCGCGGATTCACACTGCC 720
QY 721 TACTCATGCGCTTACGAGCTCGAGGGTCCCGTAAAGTCCGACCTGTGATTTGGGGCC 780
DB 721 TACTCATGCGCTTACGAGCTCGAGGGTCCCGTAAAGTCCGACCTGTGATTTGGGGCC 780
QY 781 GTCTGTGCGACGCGCGCAAGGCAACCAATTAGTCTCGAATTCATTTGCGGCAATAG 840
DB 781 GTCTGTGCGACGCGCGCAAGGCAACCAATTAGTCTCGAATTCATTTGCGGCAATAG 840
QY 841 AAACAGGTGCGGTACTGCTGATATAGCATGACAGGCGGCTGTTTCGAAGGCTCA 900
DB 841 AAACAGGTGCGGTACTGCTGATATAGCATGACAGGCGGCTGTTTCGAAGGCTCA 900
QY 901 CGACCGACCACTTACGACCAACCCGAGCTCCGCTGACGACACAGCTGTTTACTGCTG 960
DB 901 CGACCGACCACTTACGACCAACCCGAGCTCCGCTGACGACACAGCTGTTTACTGCTG 960
QY 961 GCGAACATGCGCGCTCGGTGGCGAAGAGCTGACCTAGCGGCTGACCAAGCGAGATG 1020
DB 961 GCGAACATGCGCGCTCGGTGGCGAAGAGCTGACCTAGCGGCTGACCAAGCGAGATG 1020
QY 1021 CCGTATGTGCTGAGCTTCCGCAATGCTGCGGCGGCGGCTGCGGTGCAATCCGCA 1080
DB 1021 CCGTATGTGCTGAGCTTCCGCAATGCTGCGGCGGCGGCTGCGGTGCAATCCGCA 1080
QY 1081 CTAGCCAAAGCTTTTCGACGACGAAAGGGCGTTACTGTCACAAAGGCTGGCCACCGAC 1140
DB 1081 CTAGCCAAAGCTTTTCGACGACGAAAGGGCGTTACTGTCACAAAGGCTGGCCACCGAC 1140
QY 1141 CTGGGGGTGCGGTTTCACCGGAGCGCGGAGGCTGCTGAGCTGCTGGCGGCTGTAC 1200
DB 1141 CTGGGGGTGCGGTTTCACCGGAGCGCGGAGGCTGCTGAGCTGCTGGCGGCTGTAC 1200
QY 1201 GCGGAGCACAGTGGGAGTAAGGGAAGCATGATGTCGGCCGCG 1245
DB 1201 GCGGAGCACAGTGGGAGTAAGGGAAGCATGATGTCGGCCGCG 1245

RESULT 2
LOCUS A89745 1245 bp DNA
DEFINITION Sequence 2 from Patent WO9832862.
ACCESSION A89745
VERSION A89745.1 GI:6738279
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1245)
AUTHORS Flohe, L. and Singh, M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A-30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
Source 1..1245
/db_xref="taxon:32644"
BASE COUNT 238 a 398 c 387 g 222 t
ORIGIN
Query Match 100.0%; Score 1245; DB 5; Length 1245;
Best Local Similarity 100.0%; Pred. No. 1.2e-171;
Matches 1245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTTGAGATTATCGAACTTCTTCACACTGAGAGGTACAGTATGAGAGGTAATC 60
DB 1 ATCTTGAGATTATCGAACTTCTTCACACTGAGAGGTACAGTATGAGAGGTAATC 60
QY 61 ATGGCGGTGCTATTCGACGACGACCAAAAAACAAGCAATTCATTCGGGTGGCCATC 120

```

Db 61 ATGGCGTGGTATTCCGACCGAGACCAAAAACAAGATTCAATTCCGGGTGGCCATC 120  
 QY 121 ACCCGCGCGCGGTGGGAACTAACCCGTCGTGGCCATGAGTGTCTATCCAGGAGGT 180  
 Db 121 ACCCGCGCGCGGTGGGAACTAACCCGTCGTGGCCATGAGTGTCTATCCAGGAGGT 180  
 QY 181 GCGGAGAGGGCTGGGCTATACCGAGCGGATTTCAAGGGGGGCGGCACTGCTC 240  
 Db 181 GCGGAGAGGGCTGGGCTATACCGAGCGGATTTCAAGGGGGGCGGCACTGCTC 240  
 QY 241 GGCACCGCCGACAGATGTGGGCCGAGCTGATTTATTTGCTCAAGGTCAAGAACGATA 300  
 Db 241 GGCACCGCCGACAGATGTGGGCCGAGCTGATTTATTTGCTCAAGGTCAAGAACGATA 300  
 QY 301 GCGGCGGAATACGGCGCGCTGCGACAGGGGAGATCTTGTCTGACGTCTTGCATTGGCC 360  
 Db 301 GCGGCGGAATACGGCGCGCTGCGACAGGGGAGATCTTGTCTGACGTCTTGCATTGGCC 360  
 QY 361 GCGTCACTGCTTGCACCGATGCTGTGTGATTCGGGACACGATCAATTGCTTACGAG 420  
 Db 361 GCGTCACTGCTTGCACCGATGCTGTGTGATTCGGGACACGATCAATTGCTTACGAG 420  
 QY 421 ACCGTCAGACCGCGGAGCGGCGCACTACCTCTGCTTGGCCGATGAGCGAAGTCCGGGT 480  
 Db 421 ACCGTCAGACCGCGGAGCGGCGCACTACCTCTGCTTGGCCGATGAGCGAAGTCCGGGT 480  
 QY 481 CGACTCGCGCGCGGCTTGGGCTTACCACTGATGCAACCAAGGGGGCGCGGCTGTG 540  
 Db 481 CGACTCGCGCGCGGCTTGGGCTTACCACTGATGCAACCAAGGGGGCGCGGCTGTG 540  
 QY 541 CTGATGGGCGGGGTGCCCGGCGTGAACCGGCGAGCTGTGTGATGCGGCGCGGACCC 600  
 Db 541 CTGATGGGCGGGGTGCCCGGCGTGAACCGGCGAGCTGTGTGATGCGGCGCGGACCC 600  
 QY 601 GCGGCTCAACGCAAGCCCGCATGCGCAAGGCGATGGGCGGAGCGGCTTACGCTTAC 660  
 Db 601 GCGGCTCAACGCAAGCCCGCATGCGCAAGGCGATGGGCGGAGCGGCTTACGCTTAC 660  
 QY 661 ATCAACATCGAACAACCTTGGCACTGACGCGGATCTCGGCGCGGATCAACACTGCG 720  
 Db 661 ATCAACATCGAACAACCTTGGCACTGACGCGGATCTCGGCGCGGATCAACACTGCG 720  
 QY 721 TACTCATCGGCTTACGAGCTGAGGGTCCGCTCAACGCTGCGGACCTGTGATTTGGGCC 780  
 Db 721 TACTCATCGGCTTACGAGCTGAGGGTCCGCTCAACGCTGCGGACCTGTGATTTGGGCC 780  
 QY 781 GTCTGTGTCGAGGCGGCAAGGCAACCAATTAGTCTGGAATTCATTGTGCGCATATG 840  
 Db 781 GTCTGTGTCGAGGCGGCAAGGCAACCAATTAGTCTGGAATTCATTGTGCGCATATG 840  
 QY 841 AAACAGGTGGGCTACTGCTGATATAGCATTCAGACAGGCGGCTGTTTGAAGGCTCA 900  
 Db 841 AAACAGGTGGGCTACTGCTGATATAGCATTCAGACAGGCGGCTGTTTGAAGGCTCA 900  
 QY 901 CGACGACCACTTACGAGCAACCCGAGCTTCCGCTGCAACGACAGCTTTTACTGCTG 960  
 Db 901 CGACGACCACTTACGAGCAACCCGAGCTTCCGCTGCAACGACAGCTTTTACTGCTG 960  
 QY 961 GCGACATGCGGCTGCTGAGCTTGGCGAGAGCTGACCTAC 1020  
 Db 961 GCGACATGCGGCTGCTGAGCTTGGCGAGAGCTGACCTAC 1020  
 QY 1021 GCGGATGAGGAG 1080  
 Db 1021 GCGGATGAGGAG 1080  
 QY 1081 CTAGCAAAAGTCTTTCAGACGACGAGAGGGGCTTACTGTCGAAAGGCTGCGCAC 1140  
 Db 1081 CTAGCAAAAGTCTTTCAGACGACGAGAGGGGCTTACTGTCGAAAGGCTGCGCAC 1140  
 QY 1141 CTGGGGGTGCGCTTACCGAGAGCGGCGAGGCTGCTGCTGCTGCTGCTGCTGCTTAC 1200  
 Db 1141 CTGGGGGTGCGCTTACCGAGAGCGGCGAGGCTGCTGCTGCTGCTGCTGCTGCTTAC 1200

QY 1201 GCGGAGACACAGCTGGGAGTAAGGAAGGATGATGCGGCGG 1245  
 Db 1201 GCGGAGACACAGCTGGGAGTAAGGAAGGATGATGCGGCGG 1245  
 RESULT 3  
 LOCUS A87604 1260 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 1 from Patent WO9836089.  
 ACCESSION A87604  
 VERSION A87604.1 GI:6736244  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1260)  
 AUTHORS Flohe, L. and Singh, M.  
 TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source location/Qualifiers  
 1..1260  
 /organism="unidentified"  
 /db\_xref="taxon:3264"  
 BASE COUNT 243 a 403 c 389 g 225 t  
 ORIGIN  
 Query Match 100.0%; Score 1245; DB 5; Length 1260;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-171;  
 Matches 1245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATCTTGCAATTAATGCACTTTCTTACACTGAGGCTACATGAGAGGGTAAATC 60  
 Db 16 ATCTTGCAATTAATGCACTTTCTTACACTGAGGCTACATGAGAGGGTAAATC 75  
 QY 61 ATGCGGCTGGTATCC :CGAGACCAAAACAAGCAATTC VATTGGGTGGCATC 120  
 Db 76 ATGCGGCTGGTATCC :CGAGACCAAAACAAGCAATTC VATTGGGTGGCATC 135  
 QY 121 ACCCGCGCGCGGTGGGAACTAACCCGTCGTGGCCATGAGTGTCTATCCAGGAGGT 180  
 Db 136 ACCCGCGCGCGGTGGGAACTAACCCGTCGTGGCCATGAGTGTCTATCCAGGAGGT 195  
 QY 181 GCGGAGAGGGCTGCGCTATCACGAGCGGATTTCAAGCGGCGAGCGCGCACTGCTC 240  
 Db 196 GCGGAGAGGGCTGCGCTATCACGAGCGGATTTCAAGCGGCGAGCGCGCACTGCTC 255  
 QY 241 GGCACCGCCGACAGAGTGTGGGCCGAGCTGATTTATTTGCTCAAGGTCAAGAACGATA 300  
 Db 256 GGCACCGCCGACAGAGTGTGGGCCGAGCTGATTTATTTGCTCAAGGTCAAGAACGATA 315  
 QY 301 GCGGCGGAATACGGCGCGCTGCGACAGGGGAGATCTTGTTCAGGCTTGTGATTGGCC 360  
 Db 316 GCGGCGGAATACGGCGCGCTGCGACAGGGGAGATCTTGTTCAGGCTTGTGATTGGCC 375  
 QY 361 GCGTCACTGCTTGCACCGATGCTGTGTGATTCGGGACACAGTCAATTGCTTACGAG 420  
 Db 376 GCGTCACTGCTTGCACCGATGCTGTGTGATTCGGGACACAGTCAATTGCTTACGAG 435  
 QY 421 ACCGTCAGACCGCGGAGCGGCGCACTACCTCTGCTTGGCCGATGAGCGAAGTCCGGGT 480  
 Db 436 ACCGTCAGACCGCGGAGCGGCGCACTACCTCTGCTTGGCCGATGAGCGAAGTCCGGGT 495  
 QY 481 CGACTCGCGCGCGGCTTGGCGCTTACCACTGATGCAACCAAGGGGGCGCGGCTGTG 540  
 Db 496 CGACTCGCGCGCGGCTTGGCGCTTACCACTGATGCAACCAAGGGGGCGCGGCTGTG 555  
 QY 541 CTGATGGGCGGGGTGCCCGGCGTGAACCGGCGAGCTGCTGATGATGCGGCGGACCC 600  
 Db 556 CTGATGGGCGGGGTGCCCGGCGTGAACCGGCGAGCTGCTGATGATGCGGCGGACCC 615





Db 1096 CTAGCAAAAGGTCTTTCAGACGACGAAGGGCGTTACTGTCGACAGGGGTGCGCACCGAC 1155  
 QY 1141 CTGGGGGTCCGCTTTCACCGACGCCGCGCTGCTGACTCTCGCCGCTGCTAC 1200  
 Db 1156 CTGGGGGTCCGCTTTCACCGACGCCGCGCTGCTGACTCTCGCCGCTGCTAC 1215  
 QY 1201 GCCGACACACGTCTGGAGTAAAGAGCATGATGTCGCGCGC 1245  
 Db 1216 GCCGACACACGTCTGGAGTAAAGAGCATGATGTCGCGCGC 1260

## RESULT 5

LOCUS MTALADH 2584 bp DNA BCT 01-DEC-1993  
 DEFINITION M.tuberculosis gene for L-alanine dehydrogenase.  
 ACCESSION X63069.1 GI:44565  
 VERSION X63069.1 GI:44565  
 KEYWORDS alanine dehydrogenase; extracellular; intracellular.  
 SOURCE Mycobacterium tuberculosis.  
 ORGANISM Bacteria; Firmicutes; Actinobacteriae; Actinobacteridae;  
 Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;  
 Mycobacterium.

REFERENCE 1 (bases 1 to 2584)  
 AUTHORS Andersen, A.B.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-NOV-1991) A.B. Andersen, Statens Serum Institut,  
 Artillerivej 5, DK-2300 Copenhagen S, DENMARK  
 REFERENCE 2 (bases 1 to 2584)  
 AUTHORS Andersen, A.B., Andersen, P. and Ljungqvist, L.  
 TITLE Structure and function of a 40,000-molecular-weight protein antigen  
 JOURNAL of Mycobacterium tuberculosis  
 MEDLINE Infect. Immun. 60 (6), 2317-2323 (1992)  
 FEATURES 92267644  
 source location/Qualifiers

1..2584  
 /organism="Mycobacterium tuberculosis"  
 /strain="Erdman"  
 /isolate="TMC 107, lot 9A-2"  
 /db\_xref="taxon:1773"  
 /clone\_1lb="lambda gt11"  
 /clone\_1lb="lambda AA67"  
 76..1197  
 /EC\_number="1.4.1.1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="alanine dehydrogenase"  
 /protein\_id="CAA4791.1"  
 /db\_xref="GI:44565"  
 /db\_xref="SWISS-PROT:P30234"  
 /translation="MRVGIPTETKNEPFRVAIPGVAELTRGHEVLLIGAGSEG  
 SATDADFKRAAGQIVGADQVADADLPLKVEPAAGYGRIRHGOILFTPLHLAAS  
 RACTDLDSGTSIAYEYVOTADGALPIAPAEVLAQVGYHLMRTGGGIV  
 LMGGVGPADVVVIGAGTAGYANARINATVVDINDKRLRODAEFCGRH  
 TRSSAYELEAGAKRADIVYICATVYCARAPKYSISYAHKRGVILVDIADGGC  
 FBSRPRTYDHPFAYVHDLFTFCVAMPSPVSTSYALTNTMTPLVELADHGRMAA  
 CNSPALAGLSHTEGALSERVATDLSVPTFPASVLA"

BASE COUNT 463 a 871 c 819 g 431 t  
 ORIGIN

Query Match 100.0% Score 1245; DB 1; Length 2584;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-171;  
 Matches 1245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTTGCAATTAATGCACTTTCTTCACTGAAGCGTACATATGAGAGGGGTATC 60  
 |||||||  
 Db 16 ATCTTGCAATTAATGCACTTTCTTCACTGAAGCGTACATATGAGAGGGGTATC 75  
 |||||||  
 QY 61 ATGGCGGTGATTTGCGACGAGACCAAAAGCAATTCGAATTCGGGGTGGCCATC 120  
 |||||||  
 Db 76 ATGGCGGTGATTTGCGACGAGACCAAAAGCAATTCGAATTCGGGGTGGCCATC 135  
 |||||||

QY 121 ACCCGGCGCGGCTGCGGAACTAACCCGTGCTGCGCATGAGTGTCTATCAGCAGCT 180  
 |||||||  
 Db 136 ACCCGGCGCGGCTGCGGAACTAACCCGTGCTGCGCATGAGTGTCTATCAGCAGCT 195  
 |||||||  
 QY 181 GCCGAGAGAGGCTCGGCTATACCGCAGCGGATTTCAAGGGCGGCGCAACTGTCTC 240  
 |||||||  
 Db 196 GCCGAGAGAGGCTCGGCTATACCGCAGCGGATTTCAAGGGCGGCGCAACTGTCTC 255  
 |||||||  
 QY 241 GGCACCGCGGACAGGTGTGGGCGGACGCTGATTTATGCTCAAGTGAAGAAACCGGTA 300  
 |||||||  
 Db 256 GGCACCGCGGACAGGTGTGGGCGGACGCTGATTTATGCTCAAGTGAAGAAACCGGTA 315  
 |||||||  
 QY 301 GCGGCGAATTAAGCGCGGCTGCGACGAGCGGAGATCTTGTACGCTTGTGATTTGACC 360  
 |||||||  
 Db 316 GCGGCGAATTAAGCGCGGCTGCGACGAGCGGAGATCTTGTGATTTGATTTGGCC 375  
 |||||||  
 QY 361 GCGTACGCTGCTTGCACCGCATGCTGTTGGATTCCGGACACGTCATTAATGCTACAG 420  
 |||||||  
 Db 376 GCGTACGCTGCTTGCACCGCATGCTGTTGGATTCCGGACACGTCATTAATGCTACAG 435  
 |||||||  
 QY 421 ACCGTCCAGACCGCGGAGCGGCGACTACCCCTGTTGGCCCGGATGAGGAAGTCCGCGT 480  
 |||||||  
 Db 436 ACCGTCCAGACCGCGGAGCGGCGACTACCCCTGTTGGCCCGGATGAGGAAGTCCGCGT 495  
 |||||||  
 QY 481 CGACTGCGCGCCAGGTTGGCGCTTACACACCTGATGCGAACCAGAGGAGCGCGGTGTG 540  
 |||||||  
 Db 496 CGACTGCGCGCCAGGTTGGCGCTTACACACCTGATGCGAACCAGAGGAGCGCGGTGTG 555  
 |||||||  
 QY 541 CTGATGGGGGGGTGCGCGGCTCGACCGCGGACGCTGTGTGTATGCGCGCGGAC 600  
 |||||||  
 Db 556 CTGATGGGGGGGTGCGCGGCTCGACCGCGGACGCTGTGTGTATGCGCGCGGAC 615  
 |||||||  
 QY 601 GCGGCGTAAACGAGCGCGCATCGCCCAAGCGCATGGGGCGGCGGATTAAGTAAAC 660  
 |||||||  
 Db 616 GCGGCGTAAACGAGCGCGCATCGCCCAAGCGCATGGGGCGGCGGATTAAGTAAAC 675  
 |||||||  
 QY 661 ATCAACATGCAAACTTGGCACTCGACGCGGAGTTCTGCGCGGATCCACACTGC 720  
 |||||||  
 Db 676 ATCAACATGCAAACTTGGCACTCGACGCGGAGTTCTGCGCGGATCCACACTGC 735  
 |||||||  
 QY 721 TACTCATGGGCTTAGAGCTGAGGGTGGCGTCAACGTCGCGACCTGGTATTTGGGGCC 780  
 |||||||  
 Db 736 TACTCATGGGCTTAGAGCTGAGGGTGGCGTCAACGTCGCGACCTGGTATTTGGGGCC 795  
 |||||||  
 QY 781 GTCTGTGTCGAGGCGCCAGGACACCAATATAGTCTCGAATTCATTTGTCGCGATATG 840  
 |||||||  
 Db 796 GTCTGTGTCGAGGCGCCAGGACACCAATATAGTCTCGAATTCATTTGTCGCGATATG 855  
 |||||||  
 QY 841 AAACCAAGTGGGTACTGCTGATATAGCATGACACGCGGCTGTTTGAAGGCTCA 900  
 |||||||  
 Db 856 AAACCAAGTGGGTACTGCTGATATAGCATGACACGCGGCTGTTTGAAGGCTCA 915  
 |||||||  
 QY 901 CGACGACACCTAGACACACCGGAGCTTGCGCGTGCACGACAGGCTGTTTACTGCGTG 960  
 |||||||  
 Db 916 CGACGACACCTAGACACACCGGAGCTTGCGCGTGCACGACAGGCTGTTTACTGCGTG 975  
 |||||||  
 QY 961 GCGACATCCCGCGCTCGGTGCGGAAGACGTGACCTACGCGCTGACCAACGCGAGATG 1020  
 |||||||  
 Db 976 GCGACATCCCGCGCTCGGTGCGGAAGACGTGACCTACGCGCTGACCAACGCGAGATG 1035  
 |||||||  
 QY 1021 CGGTATGTGCTGAGTGTGCGGACCATGCTGCGGGGGGCGGCGGCGGTGCAATCCGGA 1080  
 |||||||  
 Db 1036 CGGTATGTGCTGAGTGTGCGGACCATGCTGCGGGGGGCGGCGGCGGTGCAATCCGGA 1095  
 |||||||  
 QY 1081 CTAGCAAAAGGTCTTTCAGACGACGAAGGGCGTTACTGTCGACAGGGGTGCGACCGAC 1140  
 |||||||  
 Db 1096 CTAGCAAAAGGTCTTTCAGACGACGAAGGGCGTTACTGTCGACAGGGGTGCGACCGAC 1155  
 |||||||  
 QY 1141 CTGGGGGTGCTTTCACGAGCGCGGCAAGCGTGTGCTGCTGCGCGCTGCTGTTAC 1200  
 |||||||  
 Db 1156 CTGGGGGTGCTTTCACGAGCGCGGCAAGCGTGTGCTGCTGCGCGCTGCTGTTAC 1215  
 |||||||  
 QY 1201 GCCGACACACGTCTGGAGTAAAGAGCATGATGTCGCGCGC 1245

Db 1216 GCGGAGCACACGTCGGAGTAAGGAGGATGATCGCGCGC 1260

RESULT 6  
LOCUS MTW002 56414 bp DNA BCT 17-JUN-1998  
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.  
ACCESSION AL008967 AL122456  
VERSION AL008967.1 GI:3261491  
KEYWORDS  
ORGANISM Mycobacterium tuberculosis.  
SOURCE Mycobacterium tuberculosis.  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Mycobacteriaceae;  
Mycobacterium.  
1 (bases 1 to 56414)  
REFERENCE  
AUTHORS Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E., Tekle, A.F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, S., Osborn, J., Quail, M.A., Rajandream, M.A., Rogers, R., Rutter, S., Seeger, K., Skelton, S., Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and Barrett, B.G.  
TITLE Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence  
JOURNAL Nature 393 (6685), 537-544 (1998)  
MEDLINE 98295987  
REMARK Erratum: [[published erratum appears in Nature 1998 Nov 12;396(6707):190]]  
2 (bases 1 to 56414)  
Parkhill, J.  
REFERENCE  
AUTHORS Direct Submission  
TITLE Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge; CH10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
JOURNAL On Jun 27, 1998 this sequence version replaced gi:2624256.  
COMMENT Notes:  
Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.  
(URL, <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.  
Gene prediction was based on a Hidden Markov Model of TB genes implemented in tParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.  
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we have chosen an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
Location/Qualifiers  
1..56414  
/organism="Mycobacterium tuberculosis"  
/strain="H37Rv"  
/db\_xref="taxon:1773"  
misc\_feature  
1..556414  
/note="fragment designated v002. Does not represent a physical clone"  
physical clone  
1..233  
/organism="Mycobacterium tuberculosis"  
/strain="H37Rv"  
/db\_xref="taxon:1773"  
/clone="Y154"  
complement(3..527)  
/gene="recX"  
complement(3..527)  
CDS

gene  
CDS  
/gene="recX"  
/note="Rv2736c, (MTW002.01c), len: 174 aa, recX, similar to eg. RECX\_PSEAE P37860 regulatory protein recX from Pseudomonas aeruginosa (155 aa), fasta scores: opt: 161 z-score: 257.2 E(-): 3.e-07, 30.7% identity in 137 aa overlap. Overlaps and extends CDS from overlapping cosmid MTCY154.16c"  
/codon\_start=1  
/product="recX"  
/protein\_id="CA15532.1"  
/db\_xref="GI:2624257"  
/db\_xref="SPTREMBL:O33280"  
/translation="MTWSCPPPEEEDQARALCLRLTARSRAELAGADAKRGY PEDIGNRLRLAIVGLVCDPPEAEQVQSRANAAKRAELAEHAGVDVYIT VLGIDGAEGRRAEKLVRARLRREVIDIDGTDEARSRVLVMLARIGQLACEV VLAELAEERERRV"  
complement(493..2865)  
/gene="recA"  
complement(493..2865)  
/gene="recA"  
/note="Rv2737c, (MTW002.02c), len: 790 aa, recA, identical to recA\_MYCTU P26345 recA protein (790 aa). Contains self-splicing protein element (intron) from 2294 to 969 (c), similar to intron II from TR:E332317 (EMBL:Y13030) DNA-directed DNA polymerase (EC 2.7.7.7) from Thermococcus sp. (1829 aa), fasta scores: opt: 81 z-score: 235.2 E(-): 6e-06, 24.6% identity in 183 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop), PS00321 recA signature, and PS00881 protein splicing signature. See Davis et al, (1992) Cell 71(2):201-210"  
/codon\_start=1  
/product="recA"  
/protein\_id="CA15533.1"  
/db\_xref="GI:2624258"  
/db\_xref="SWISS-PROT:P26345"  
/translation="MTQTPPREAKLELAVAQIEKSYKSGVRLGDEAROPTISVPTG SIALDVALLGIGLPRGRVIEIYGPSSGKTTLALHVAANAAGVAALFADAHADP DYARKLGVDDLSLVSDPDGEOLAEADMLIRSGALDIYIVDSVALVPRALBEEM GDSHVGLAOLMAGLAKMKGALNNGTTFINOLBDRKIGVAPGSPETTRGAKAF YASRMVRRVETLKDGTNAVGNTRKRVKNCCLAGCTIFPPTVGTTRIDVDYG RKLHVVAANKDGLHARPVASWFDQSTRIQLRLAGIAIVATPDHVLVEYGRA AGELRKGRDVAQPRDFGDSAPIPDARHRLGLYLIGDRDMDVGKPPINEINOR ALIDVYRIATLGCAGAPOGRIASLAHPRGSRNGVADICQAGIYKLAKEITPN WFEPPDIADIVGILGELGSPDGWSREOTGALRVGTTSSQLAHQIMLLIRGV GSTYRDVDPTRKPSIVNGRIOSKRVQFEVRSIGDMNTAPAEVPMGPRGALIO AIPRATQRRRGSAATYLAEMTDVANTIDERGVTQAEAAATIGVAGSDPRGMQV LGASRLRRDVAQADALDLDKFLHDMIAEELRSVIREVLPTRARFTDLEVEHLTV VAGSVVHNGSPPEKQAEFDILYKGISRGSLLIDMGVDGLIRKSGAMFTYEGEOLG OGEKMANPFLVENADVADIEIKKIKELGIGAVTDPDSNDGLVPAPVF"  
complement(790..2112)  
/gene="recA"  
/note="recA intein"  
complement(790..807)  
/gene="recA"  
/note="PS00881 protein splicing signature"  
complement(2197..2223)  
/gene="recA"  
/note="PS00321 recA signature"  
complement(2644..2667)  
/gene="recA"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
complement(2873..2877)  
/note="possible RBS upstream of Rv2737c"  
complement(3247..3453)  
/gene="Rv2738c"  
complement(3247..3453)  
/gene="Rv2738c"  
/note="Rv2738c, (MTW002.03c), len: 68 aa, unknown. N-terminus is highly similar to the N-terminus of the upstream ORFMTW002.07c (78.4% identity in 37 aa overlap); also similar to AL020958|SC4H8\_5 streptomycetes coellicolor

cosmid 4H8 (64aa) opt: 185 z-score: 283.5 E(): 2.9e-08;  
39.7% identity in 63 aa overlap  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein Rv2738c"  
/protein\_id="CA15534.1"  
/db\_xref="GI:2624260"  
/db\_xref="SPTREMBL:O33281"  
/translation="MLAGVRLTEPHERVALHFGAAYSSVLDHVLTFPGRSAQAQAI  
EDGVEPDRVALCADFVPRDRM"  
/complement(3464..4630)  
/gene="Rv2739c"  
/note="Rv2739c. (MTV002.04c), len: 388 aa. Probable  
transferase, similar to eg. TR-Q051560 (EMBL:L28170)  
Pseudomonas aeruginosa thiamosyl transferase (426 aa),  
fasta scores: opt: 178 z-score: 226.3 E(): 1.9e-05, 25.98  
identity in 425 aa overlap. Equivalent to Mycobacterium  
leprae protein MLCB3.02c (392 aa); fasta scores  
gpl294723MLCB3.2 Mycobacterium leprae cosmid B33 opt:  
2112 z-score: 2364.5 E(): 0; 80.98 identity in 388 aa  
overlap  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein Rv2739c"  
/protein\_id="CA15535.1"  
/db\_xref="GI:2624261"  
/db\_xref="SPTREMBL:O33282"  
/translation="MRVAVVAGPPDGHSPFALICQRRRA..DTPLTFVYEMLEAR  
AAGIDAVELDLAATRDLDGARHHRRAQAVLVNRLALPELVSQVITACGG  
MAELLAGIPWELNPHLPKLPKGLPPIGSLAAGTGLRDLATMRLALGRSMRG  
LRORAVAVEIGLPAKDPGPRRLIALIPALVEPRDPAAVAVVGRHPEPTDRA  
IPAGGPPVAVVAPSTALGTAGLEVALQSLPGETVSGSRLVYRSLGADLPVPM  
AVAGISQAEITRADIVICGGHGMVAKTLAGAPVAVVGGDDQMEIARVYRQSS  
AVLIRPLADALVAVNEVLSPPRFRARRAASVAGAAPVAVCHDALALAG"  
4674..5123  
/gene="Rv2740"  
4674..5123  
/gene="Rv2740"  
/note="Rv2740. (MTV002.05), len: 149 aa. unknown, similar  
to M. leprae hypothetical protein TR-Q049850 (EMBL:294723)  
MLCB3.03 (B2235\_F3\_140) (178 aa), fasta scores: opt: 498  
z-score: 538.9 E(): 7.3e-23, 51.68 identity in 161 aa  
overlap  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein Rv2740"  
/protein\_id="CA15536.1"  
/db\_xref="GI:2624262"  
/db\_xref="SPTREMBL:O33283"  
/translation="MAELTESPPTETEARAVEAFNLQNEPDTVDALGDDL  
YVENGFIRIGRRTATLRMGQRFVEYKIHIGADGAVALERTDALIPLRY  
QWVGVFEVDGRTLRWDYEDVDMFKLLRGLVALVPSLTKATL"  
5355..6932  
/gene="PE\_PGRS"  
5355..6932  
/gene="PE\_PGRS"  
/note="Rv2741. (MTV002.06), len: 525 aa; Member of M.  
tuberculosis PE\_PGRS subfamily; similar to many eg.

Query Match 98.1%; Score 1221.4; DB 1; Length 56414;  
Best local Similarity 99.4%; Pred. No. 1.4e-168;  
Matches 1238; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

QY 1 ATCTTCAGATTATGACCTTCTTCACTGACGTCGACATATCGAGAGGGGTAATC 60  
DB 38201 ATCTTCAGATTATGACCTTCTTCACTGACGTCGACATATCGAGAGGGGTAATC 38260  
QY 61 ATGCGGCTGCTATTCGACGACGACCAAAACAGCAATTCATTCGGGTGGCATC 120  
DB 38261 ATGCGGCTGCTATTCGACGACGACCAAAACAGCAATTCATTCGGGTGGCATC 38314

QY 121 ACCCGCGCGCGCTGCGGGAAGTAACCCGTCGTGGCATGAGGTGCTCATCCAGGAGGT 180  
DB 38315 ACCCGCGCGCGCTGCGGGAAGTAACCCGTCGTGGCATGAGGTGCTCATCCAGGAGGT 38374  
QY 181 GCGGAGAGGAGGCTCGGCTATCCAGCAGCGGATTTCAAGGGGCGAGCGCGCAATGCTC 240  
DB 38375 GCGGAGAGGAGGCTCGGCTATCCAGCAGCGGATTTCAAGGGGCGAGCGCGCAATGCTC 38434  
QY 241 GCGAGCGCGCGCGAGGCTGCGGCGCGAGCGGATTTGCTCAAGGTCAAGAACCGGATA 300  
DB 38435 GCGAGCGCGCGCGAGGCTGCGGCGCGAGCGGATTTGCTCAAGGTCAAGAACCGGATA 38494  
QY 301 GCGGCGGAATATAGCGCGCGCTGCGACACAGCGGAGATCTTGTTCAGCTTCTTGCATTGGCC 360  
DB 38495 GCGGCGGAATATAGCGCGCGCTGCGACACAGCGGAGATCTTGTTCAGCTTCTTGCATTGGCC 38554  
QY 361 GCGTCAGCTGCTTGCACCGATGCGTGTGATTCGCGGACACAGCTCAATGCTTACGAG 420  
DB 38555 GCGTCAGCTGCTTGCACCGATGCGTGTGATTCGCGGACACAGCTCAATGCTTACGAG 38614  
QY 421 ACCGTCCAGACCGCGCGAGCGGCGGACCTACCCCTGCTTGGCCGATGAGCGAAGTGGCGGT 480  
DB 38615 ACCGTCCAGACCGCGCGAGCGGCGGACCTACCCCTGCTTGGCCGATGAGCGAAGTGGCGGT 38674  
QY 481 CGACTCGCGCGCGCGAGTGTGGCGCTTACACCTGATGCGAACCCAGGGGCGCGGTTGTG 540  
DB 38675 CGACTCGCGCGCGCGAGTGTGGCGCTTACACCTGATGCGAACCCAGGGGCGCGGTTGTG 38734  
QY 541 CTGATGGCGGCGGCGTCCCGCGCTGGAACCGCGGACGCTGCTGATGCGCGCGCGGAC 600  
DB 38735 CTGATGGCGGCGGCGTCCCGCGCTGGAACCGCGGACGCTGCTGATGCGCGCGGAC 38794  
QY 601 GCGCGCTACAACGAGCAGCGCGCGATCGCGCAAGCGCATGGCGCGAGACCGTTACGTTAGAC 660  
DB 38795 GCGCGCTACAACGAGCAGCGCGCGATCGCGCAAGCGCATGGCGCGAGACCGTTACGTTAGAC 38854  
QY 661 ATCAACATGCAAACTTGGGCAACTGACGCGGAGTTCGCGCGCGGATTCACACTCGC 720  
DB 38855 ATCAACATGCAAACTTGGGCAACTGACGCGGAGTTCGCGCGCGGATTCACACTCGC 38914  
QY 721 TACGTCATGCGCGCTGAGGCTGAGGAGTGGCGTCAAGGCGGCGCGGATTTGGGGCC 780  
DB 38915 TACGTCATGCGCGCTGAGGCTGAGGAGTGGCGTCAAGGCGGCGCGGATTTGGGGCC 38974  
QY 781 GTCTGTGTGCGGCGCGCGCAAGGACCGCAATTAAGTCTGATTCCTGCTGCGCATATG 840  
DB 38975 GTCTGTGTGCGGCGCGCGCAAGGACCGCAATTAAGTCTGATTCCTGCTGCGCATATG 39034  
QY 841 AAACAGGTGCGGTAAGTGTGATATACCATGACACAGGCGGCTGTTGGAAGGCTCA 900  
DB 39035 AAACAGGTGCGGTAAGTGTGATATACCATGACACAGGCGGCTGTTGGAAGGCTCA 39094  
QY 901 CGACGACGACCTGAGACACCGCGCGTGGCGTGGACGACGAGGCTTTTACTGGGTG 960  
DB 39095 CGACGACGACCTGAGACACCGCGCGTGGCGTGGACGACGAGGCTTTTACTGGGTG 39154  
QY 961 GCGAATCATGCGCGCGCTGCTGCTGCGAAGAGCTGAGAGCTGAGAGGAGGAGGAGGAG 1020  
DB 39155 GCGAATCATGCGCGCGCTGCTGCTGCGAAGAGCTGAGAGCTGAGAGGAGGAGGAGGAG 39214  
QY 1021 CCGTATGTGCTGAGACTTGGCGACCATGCTGGGCGGCGGCGTCCGAGTCAATCGGCA 1080  
DB 39215 CCGTATGTGCTGAGACTTGGCGACCATGCTGGGCGGCGGCGTCCGAGTCAATCGGCA 39274  
QY 1081 CTAGCCAAAGGCTTTCAGCAGCAGCAAGGCGGCTTACGTCCGAAGGAGGCGCACAGAC 1140  
DB 39275 CTAGCCAAAGGCTTTCAGCAGCAGCAAGGCGGCTTACGTCCGAAGGAGGCGCACAGAC 39334  
QY 1141 CTGGGCGGTGCGTTCACGAGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTTAC 1200  
DB 39335 CTGGGCGGTGCGTTCACGAGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTTAC 39394  
QY 1201 GCCGAGCACACGCTGGGAGTAAAGGAGGAGGATGATGTCGCGCGC 1245

Db 39395 GCCGAGCAGACGTGGGAGTAAAGGAGCGATGATGTGCGCGCG 39439

RESULT 7  
LOCUS AB7613 1236 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 10 from Patent WO9836089.  
ACCESSION AB7613  
VERSION AB7613.1 GI:6736253  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified  
unclassified.

REFERENCE 1 (bases 1 to 1236)  
AUTHORS Flohe, L. and Singh, M.  
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
FEATURES  
source 1..1236  
Location/Qualifiers  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 236 a 395 c 385 g 220 t  
ORIGIN

Query Match 98.0%; Score 1220; DB 5; Length 1236;  
Best Local Similarity 99.5%; Pred. No. 5.1e-168;  
Matches 1236; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

Db 1 ATCTTGAGATTATCGAATTCTTCTCACAAGGAGTACAGTATGAGAGGGTAATC 60  
1 ATCTTGAGATTATCGAATTCTTCTCACAAGGAGTACAGTATGAGAGGGTAATC 60  
QY 61 ATGGCGCTGCTATTCGACCGAGACCAAAACAACGAATTCATTCGGGTGCCATC 120  
61 ATGGCGCTGCTATTCGACCGAGACCAAAACAACGAATTCATTCGGGTGCCATC 114  
Db 61 ATGGCGCTGCTATTCGACCGAGACCAAAACAACGAATTCATTCGGGTGCCATC 114  
QY 121 ACCCGGGCGGCTGCGGGAAGTAACCGGTGCGGCAATGAGTCTCATCCAGGAGT 180  
121 ACCCGGGCGGCTGCGGGAAGTAACCGGTGCGGCAATGAGTCTCATCCAGGAGT 180  
Db 115 ACCCGGGCGGCTGCGGGAAGTAACCGGTGCGGCAATGAGTCTCATCCAGGAGT 174  
115 ACCCGGGCGGCTGCGGGAAGTAACCGGTGCGGCAATGAGTCTCATCCAGGAGT 174  
QY 181 GCCGAGAGGGCTCGGCTATCACCGAGCGGATTCAGAGCGGCGCAACTGCTC 240  
181 GCCGAGAGGGCTCGGCTATCACCGAGCGGATTCAGAGCGGCGCAACTGCTC 240  
Db 175 GCCGAGAGGGCTCGGCTATCACCGAGCGGATTCAGAGCGGCGCAACTGCTC 234  
175 GCCGAGAGGGCTCGGCTATCACCGAGCGGATTCAGAGCGGCGCAACTGCTC 234  
QY 241 GGCACCGCCGACGAGGTGTGGGCGGAGCTATTTATGCTCAAGGTCAAGAACCGATA 300  
241 GGCACCGCCGACGAGGTGTGGGCGGAGCTATTTATGCTCAAGGTCAAGAACCGATA 300  
Db 235 GGCACCGCCGACGAGGTGTGGGCGGAGCTATTTATGCTCAAGGTCAAGAACCGATA 294  
235 GGCACCGCCGACGAGGTGTGGGCGGAGCTATTTATGCTCAAGGTCAAGAACCGATA 294  
QY 301 GCGGCGGAATACGGCGGCTCGACAGCGGAGATCTTGTTCAGTCTTCTCATTTGGCC 360  
301 GCGGCGGAATACGGCGGCTCGACAGCGGAGATCTTGTTCAGTCTTCTCATTTGGCC 360  
Db 295 GCGGCGGAATACGGCGGCTCGACAGCGGAGATCTTGTTCAGTCTTCTCATTTGGCC 354  
295 GCGGCGGAATACGGCGGCTCGACAGCGGAGATCTTGTTCAGTCTTCTCATTTGGCC 354  
QY 361 GCGTACAGTCTTGCACCGATGCGTGTGTGATTCCGGACACAGTCAATTGCTTACGAG 420  
361 GCGTACAGTCTTGCACCGATGCGTGTGTGATTCCGGACACAGTCAATTGCTTACGAG 420  
Db 355 GCGTACAGTCTTGCACCGATGCGTGTGTGATTCCGGACACAGTCAATTGCTTACGAG 414  
355 GCGTACAGTCTTGCACCGATGCGTGTGTGATTCCGGACACAGTCAATTGCTTACGAG 414  
QY 421 ACCGTCCAGACCGCGGAGCGGCACTACCTGCTTGGCCCCGATGAGCGAAGTCCCGGT 480  
421 ACCGTCCAGACCGCGGAGCGGCACTACCTGCTTGGCCCCGATGAGCGAAGTCCCGGT 480  
Db 415 ACCGTCCAGACCGCGGAGCGGCACTACCTGCTTGGCCCCGATGAGCGAAGTCCCGGT 474  
415 ACCGTCCAGACCGCGGAGCGGCACTACCTGCTTGGCCCCGATGAGCGAAGTCCCGGT 474  
QY 481 CGACTCGCGCGCGAGGTTGGGCTTACCACTGATGGAAACCAAGGGGGCGCGGGTGTG 540  
481 CGACTCGCGCGCGAGGTTGGGCTTACCACTGATGGAAACCAAGGGGGCGCGGGTGTG 540  
Db 475 CGACTCGCGCGCGAGGTTGGGCTTACCACTGATGGAAACCAAGGGGGCGCGGGTGTG 534  
475 CGACTCGCGCGCGAGGTTGGGCTTACCACTGATGGAAACCAAGGGGGCGCGGGTGTG 534  
QY 541 CTGATGGCGGGGTGCCCGGCGTGAACCGGCGACGTCGTGTGATCGGCGCGGACCC 600  
541 CTGATGGCGGGGTGCCCGGCGTGAACCGGCGACGTCGTGTGATCGGCGCGGACCC 600  
Db 535 CTGATGGCGGGGTGCCCGGCGTGAACCGGCGACGTCGTGTGATCGGCGCGGACCC 594  
535 CTGATGGCGGGGTGCCCGGCGTGAACCGGCGACGTCGTGTGATCGGCGCGGACCC 594  
QY 601 GCGGCGTCAACGACCGCGCGATGCCAACCGCATGGCAGGACCGTTACGCTTACGAC 660  
601 GCGGCGTCAACGACCGCGCGATGCCAACCGCATGGCAGGACCGTTACGCTTACGAC 660

Db 595 GCCGCGTACACGACCGCCGATCGCCAAAGCGATGGCGCGACCGTTACGTTACAG 654  
QY 661 ATCAACATGACAACTTGGCACTGACAGCGCGAGTCTGCGGCGGATTCACACTGCG 720  
661 ATCAACATGACAACTTGGCACTGACAGCGCGAGTCTGCGGCGGATTCACACTGCG 714  
Db 655 ATCAACATGACAACTTGGCACTGACAGCGCGAGTCTGCGGCGGATTCACACTGCG 714  
QY 721 TACTCATCGGCTTACGAGCTCGAGGGTGGCGTCAACAGTCCGACCTGATGGGGCC 780  
721 TACTCATCGGCTTACGAGCTCGAGGGTGGCGTCAACAGTCCGACCTGATGGGGCC 780  
Db 715 TACTCATCGGCTTACGAGCTCGAGGGTGGCGTCAACAGTCCGACCTGATGGGGCC 774  
715 TACTCATCGGCTTACGAGCTCGAGGGTGGCGTCAACAGTCCGACCTGATGGGGCC 774  
QY 781 GTCCTGTCGACAGCGCCCAAGGACCCAAATTAGTCTCAATTCACTTTCGCGCATATG 840  
781 GTCCTGTCGACAGCGCCCAAGGACCCAAATTAGTCTCAATTCACTTTCGCGCATATG 840  
Db 775 GTCCTGTCGACAGCGCCCAAGGACCCAAATTAGTCTCAATTCACTTTCGCGCATATG 834  
775 GTCCTGTCGACAGCGCCCAAGGACCCAAATTAGTCTCAATTCACTTTCGCGCATATG 834  
QY 841 AAACCAAGTCCGCTACTGCTGATATAGCCATGACACGAGGGGCTGTTCCAGAGCTCA 900  
841 AAACCAAGTCCGCTACTGCTGATATAGCCATGACACGAGGGGCTGTTCCAGAGCTCA 900  
Db 835 AAACCAAGTCCGCTACTGCTGATATAGCCATGACACGAGGGGCTGTTCCAGAGCTCA 894  
835 AAACCAAGTCCGCTACTGCTGATATAGCCATGACACGAGGGGCTGTTCCAGAGCTCA 894  
QY 901 CGACCGACACCTACGACACCCGAGCTTGGCGGTGCACAGACGCTGTTTACTGCTG 960  
901 CGACCGACACCTACGACACCCGAGCTTGGCGGTGCACAGACGCTGTTTACTGCTG 960  
Db 895 CGACCGACACCTACGACACCCGAGCTTGGCGGTGCACAGACGCTGTTTACTGCTG 954  
895 CGACCGACACCTACGACACCCGAGCTTGGCGGTGCACAGACGCTGTTTACTGCTG 954  
QY 961 GCGAAGATGCGCGCTCGGTGCGGAGAGCTGACCTACGCGCTGACCAACGCGAGATG 1020  
961 GCGAAGATGCGCGCTCGGTGCGGAGAGCTGACCTACGCGCTGACCAACGCGAGATG 1020  
Db 955 GCGAAGATGCGCGCTCGGTGCGGAGAGCTGACCTACGCGCTGACCAACGCGAGATG 1014  
955 GCGAAGATGCGCGCTCGGTGCGGAGAGCTGACCTACGCGCTGACCAACGCGAGATG 1014  
QY 1021 CCGTATGTGCTCAGCTTGGCCGACATGGCTGGCGGGCGGCTGCGGCTGCAATCGGCA 1080  
1021 CCGTATGTGCTCAGCTTGGCCGACATGGCTGGCGGGCGGCTGCGGCTGCAATCGGCA 1080  
Db 1015 CCGTATGTGCTCAGCTTGGCCGACATGGCTGGCGGGCGGCTGCGGCTGCAATCGGCA 1074  
1015 CCGTATGTGCTCAGCTTGGCCGACATGGCTGGCGGGCGGCTGCGGCTGCAATCGGCA 1074  
QY 1081 CTAGCCAAAGTCTTTCGACGACGCAAGAGGCGCTTACTCTCCGAAGGGTGGGACACGAC 1140  
1081 CTAGCCAAAGTCTTTCGACGACGCAAGAGGCGCTTACTCTCCGAAGGGTGGGACACGAC 1140  
Db 1075 CTAGCCAAAGTCTTTCGACGACGCAAGAGGCGCTTACTCTCCGAAGGGTGGGACACGAC 1134  
1075 CTAGCCAAAGTCTTTCGACGACGCAAGAGGCGCTTACTCTCCGAAGGGTGGGACACGAC 1134  
QY 1141 CTGGGGGTGCGCTTTCACGACGACCGCCAGCGTGTGCTGCTGACTCTGCGCGCTGTTAC 1200  
1141 CTGGGGGTGCGCTTTCACGACGACCGCCAGCGTGTGCTGCTGACTCTGCGCGCTGTTAC 1200  
Db 1135 CTGGGGGTGCGCTTTCACGACGACCGCCAGCGTGTGCTGCTGACTCTGCGCGCTGTTAC 1194  
1135 CTGGGGGTGCGCTTTCACGACGACCGCCAGCGTGTGCTGCTGACTCTGCGCGCTGTTAC 1194  
QY 1201 GCGGAGACACGTCGGGAGTAAGGAAGCATGATCTCGGCC 1242  
1201 GCGGAGACACGTCGGGAGTAAGGAAGCATGATCTCGGCC 1242  
Db 1195 GCGGAGACACGTCGGGAGTAAGGAAGCATGATCTCGGCC 1236  
1195 GCGGAGACACGTCGGGAGTAAGGAAGCATGATCTCGGCC 1236

RESULT 8  
LOCUS AB9753 1236 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 10 from Patent WO9832862.  
ACCESSION AB9753  
VERSION AB9753.1 GI:6748287  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified  
unclassified.

REFERENCE 1 (bases 1 to 1236)  
AUTHORS Flohe, L. and Singh, M.  
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
FEATURES  
source 1..1236  
Location/Qualifiers  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 236 a 395 c 385 g 220 t  
ORIGIN

Query Match 98.0%; Score 1220; DB 5; Length 1236;  
Best Local Similarity 99.5%; Pred. No. 5.1e-168;  
Matches 1236; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

QY 1 ATCTTGAGATTATCGAATTCTTCTCACAAGGAGTACAGTATGAGAGGGTAATC 60  
1 ATCTTGAGATTATCGAATTCTTCTCACAAGGAGTACAGTATGAGAGGGTAATC 60  
Db 1 ATCTTGAGATTATCGAATTCTTCTCACAAGGAGTACAGTATGAGAGGGTAATC 60

61 ATGCGCGTGGTATTCGACGAGACCAAAACAAATTCATTCGGGGTGGCATC 120  
 61 ATGCGCGTGGTATTCGACGAGACCAAAACAAACG-----AATTCGGGGTGGCATC 114  
 121 ACCCGGGCGGGTGGGAACTAACCGGTGTCGTCATGAGTGTCTCATCCAGGAGT 180  
 115 ACCCGGGCGGGTGGGAACTAACCGGTGTCGTCATGAGTGTCTCATCCAGGAGT 174  
 181 GCGGAGAGGCGTGGCTATTCACCGAGCGGATTTCAAGCGGCGAGCGCGCACTGGTC 240  
 175 GCGGAGAGGCGTGGCTATTCACCGAGCGGATTTCAAGCGGCGAGCGCGCACTGGTC 234  
 241 GCGACCGCGGACGAGTGTGGCGCGAGCTGATTTATGCTCAAGGTCAAGAACGATA 300  
 235 GCGACCGCGGACGAGTGTGGCGCGAGCTGATTTATGCTCAAGGTCAAGAACGATA 294  
 301 GCGGCGGAATACGGCGCGCTGCGACACGGGCGAGATCTTGTCACTTGTTCATTTGGCC 360  
 295 GCGGCGGAATACGGCGCGCTGCGACACGGGCGAGATCTTGTCACTTGTTCATTTGGCC 354  
 361 GCGTACGCTGTGACGATGCTGTGTGATTCGGGCACGTCATTTGGCTTCGAG 420  
 355 GCGTACGCTGTGACGATGCTGTGTGATTCGAGACGACGTCATTTGGCTTCGAG 414  
 421 ACCGTCAACGCGCGAGCGGCGCATACCCCTGCTTGGCCCGATGAGCAAGTCGCGGT 480  
 415 ACCGTCAACGCGCGCGAGCGGCGCATACCCCTGCTTGGCCCGATGAGCAAGTCGCGGT 474  
 481 CGACTCGCGCGCGAGTGGCGCTTACACCTGATGCGAAACCAAGGGGCGCGGTGTG 540  
 475 CGACTCGCGCGCGAGTGGCGCTTACACCTGATGCGAAACCAAGGGGCGCGGTGTG 534  
 541 CTGATGGGGGGTGGCGCGCGTTCGAACCGCGGAGTGTGTATTCGGCGCGCGCAC 600  
 535 CTGATGGGGGGTGGCGCGCGTTCGAACCGCGGAGTGTGTATTCGGCGCGCGCAC 594  
 601 GCGGCGTACACGACGCGCGCATCGCCAAAGCGATGGGCGCGAGCTTACGCTTACAG 660  
 595 GCGGCGTACACGACGCGCGCATCGCCAAAGCGATGGGCGCGAGCTTACGCTTACAG 654  
 661 ATCAACATGACAACTTGGGCACTCGAGCGGAGTGTGGCGCGGATTCACACTGCG 720  
 655 ATCAACATGACAACTTGGGCACTCGAGCGGAGTGTGGCGCGGATTCACACTGCG 714  
 721 TACTATCGGCTTACGAGTTCGAGGCTGCGTAAACGTCGACCTGTGATTTGGGCG 780  
 715 TACTATCGGCTTACGAGTTCGAGGCTGCGTAAACGTCGACCTGTGATTTGGGCG 774  
 781 GTCCTGTGCGAGGCGCGCAAGGCAACCAATTTAGTCTGCAATTCACCTTGTGCGCATATG 840  
 775 GTCCTGTGCGAGGCGCGCAAGGCAACCAATTTAGTCTGCAATTCACCTTGTGCGCATATG 834  
 841 AAACGAGTGGTACTGTGATATGCAATGACAGGCGGCTGTTCGAAGGCTCA 900  
 835 AAACGAGTGGTACTGTGATATGCAATGACAGGCGGCTGTTCGAAGGCTCA 894  
 901 GACGAGCACTACGACACCGAGTTCGCGTGCAGACGAGCTGTTCGAGTTCGAGT 960  
 895 GACGAGCACTACGACACCGAGTTCGCGTGCAGACGAGCTGTTCGAGTTCGAGT 954  
 961 GCGAATATCGCGCTGCGTGCAGAGTTCGAGTTCGAGTTCGAGTTCGAGT 1020  
 955 GCGAATATCGCGCTGCGTGCAGAGTTCGAGTTCGAGTTCGAGTTCGAGT 1014  
 1021 CCGTATGCTGAGCTTGGCGACATGCTGCGGCGGCGCTGCGGCTGCAATCCGCA 1080  
 1015 CCGTATGCTGAGCTTGGCGACATGCTGCGGCGGCGCTGCGGCTGCAATCCGCA 1074  
 1081 CTAGCCAAAGGCTTTCGAGCAGAGGAGGCTTACTGCGAAGGGTGGCGACGAG 1140  
 1075 CTAGCCAAAGGCTTTCGAGCAGAGGAGGCTTACTGCGAAGGGTGGCGACGAG 1134

1141 CTGGGGTCCCTTCACGAGCCCGCAAGCGTGGCTGACTCTCGGCGCGCTGATC 1200  
 1135 CTGGGGTCCCTTCACGAGCCCGCAAGCGTGGCTGACTCTCGGCGCGCTGATC 1194  
 1201 GCCAGACACGCTGGGAGTAAGGAACGATGATGCGCC 1242  
 1195 GCCAGACACGCTGGGAGTAAGGAACGATGATGCGCC 1236  
 RESULT 9  
 1237 bp DNA PAT 22-JAN-2000  
 LOCUS A87607  
 DEFINITION Sequence 4 from Patent WO9836089.  
 ACCESSION A87607  
 VERSION A87607.1 GI:6736247  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 unidentifed.  
 unidentifed.  
 unclassified.  
 unclassified.  
 REFERENCE  
 1 (bases 1 to 1237)  
 AUTHORS Flohe L. and Singh M.  
 TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source  
 1..1237  
 /organism="unidentifed"  
 /db\_xref="taxon:32644"  
 BASE COUNT 236 a 394 c 386 g 221 t  
 ORIGIN  
 Query Match 97.9%; Score 1219.4; DB 5; Length 1237;  
 Best Local Similarity 99.4%; Pred. No. 6.3e-166;  
 Matches 1236; Conservative 0; Mismatches 1; Indels 6; Gaps 1;  
 1 ATCTTGACATTAATGAACTTCTTCACTGAGAGCGATGATGAGAGGTAATC 60  
 1 ATCTTGACATTAATGAACTTCTTCACTGAGAGCGATGATGAGAGGTAATC 60  
 61 ATGCGGTGCTGATTCG...CGAGACCAAAACAAAGATTCATTCGCGGTGCGATC 120  
 61 ATGCGGTGCTGATTCG...CGAGACCAAAACAAAGATTCATTCGCGGTGCGATC 120  
 121 ACCCGCGCGGCGTGGCGAATACCGCGTGTGCGCATGAGTGTCTCAACAGCAGT 180  
 115 ACCCGCGCGGCGTGGCGAATACCGCGTGTGCGCATGAGTGTCTCAACAGCAGT 174  
 181 GCGGAGAGGCGTGGCGAATACCGCGTGTGCGCATGAGTGTCTCAACAGCAGT 240  
 175 GCGGAGAGGCGTGGCGAATACCGCGTGTGCGCATGAGTGTCTCAACAGCAGT 234  
 241 GGCACCGCGGACGAGTGTGGCGCGAGCTGATTTATGCTCAAGTCAAGAACGATA 300  
 235 GGCACCGCGGACGAGTGTGGCGCGAGCTGATTTATGCTCAAGTCAAGAACGATA 294  
 301 GCGGCGGAATACGGCGCGCTGCGACACGGCGAGATCTTGTTCAGTTCGATTTGGCC 360  
 295 GCGGCGGAATACGGCGCGCTGCGACACGGCGAGATCTTGTTCAGTTCGATTTGGCC 354  
 361 GCGTACGCTGTGACGATGCTGTGTGATTCGGGCGACGCTCAATTTGGCGAG 420  
 355 GCGTACGCTGTGACGATGCTGTGTGATTCGGGCGACGCTCAATTTGGCGAG 414  
 421 ACCGTCAACGCGCGAGCGGCGCATACCCCTGCTTGGCCCGATGAGCAAGTCGCGGT 480  
 415 ACCGTCAACGCGCGAGCGGCGCATACCCCTGCTTGGCCCGATGAGCAAGTCGCGGT 474  
 481 CGACTCGCGCGCGAGTGGCGCTTACACCTGATGCGAAACCAAGGGGCGCGGTGTG 540  
 475 CGACTCGCGCGCGAGTGGCGCTTACACCTGATGCGAAACCAAGGGGCGCGGTGTG 534  
 541 CTGATGGGGGGTGGCGCGCGTTCGAACCGGCGAGCGTGTGATGAGGCGCGGCGAC 600



```

|||||
Db 535 CTGATGGGGGGGGTGGCCGGCTCCAGACCGGCGGAGCGTGTGTATGGGGCGGGACC 594
OY 601 GCGGGCTACAGCAGCAGCCCGCATCGCCAAAGGCGATGGGCGCAGCGTTACGGTTTAAAC 660
Db 595 GCGGGCTACAGCAGCAGCCCGCATCGCCAAAGGCGATGGGCGCAGCGTTTAAAC 654
OY 661 ATCAACATGACAACTTGGGCACTCGAGCGCGGATTTGGGGCGGATTCACACTGC 720
Db 655 ATCAACATGACAACTTGGGCACTCGAGCGCGGATTTGGGGCGGATTCACACTGC 714
OY 721 TACTCATCGGCTACGAGCTCGAGGGTCCGTCAAAGTCCGACCTGTATGGGGCC 780
Db 715 TACTCATCGGCTACGAGCTCGAGGGTCCGTCAAAGTCCGACCTGTATGGGGCC 774
OY 781 GTCTGTGTCGAGGGCGGCAAGGCGCAATTAAGTGTGAATTCCTTGGCGCATATG 840
Db 775 GTCTGTGTCGAGGGCGGCAAGGCGCAATTAAGTGTGAATTCCTTGGCGCATATG 834
OY 841 AAACGAGTGGGTACTGTGTATATAGCATAGCAGCAGGCGGGCTGTTCGAAGGCTCA 900
Db 835 AAACGAGTGGGTACTGTGTATATAGCATAGCAGCAGGCGGGCTGTTCGAAGGCTCA 894
OY 901 CGACGACCACTACGACCAACCGGAGCTGCGCGTGCAGAGACAGGTTTAAAGCGTG 960
Db 895 CGACGACCACTACGACCAACCGGAGCTGCGCGTGCAGAGTGCATCTGTTACTCTG 954
OY 961 GCAACATGCGCGCTCGGTGCGGAGAGAGCTGAGCTAGCGGCTGACCAAGCGAGATG 1020
Db 955 GCAACATGCGCGCTCGGTGCGGAGAGAGCTGAGCTAGCGGCTGACCAAGCGAGATG 1014
OY 1021 CGGTATGTGCTGAGCTTGGCGACATGCTGCGGGGGCGGCTGCGGTGCAATCCGGA 1080
Db 1015 CCTATGTGCTGAGCTTGGCGACATGCTGCGGGGGCGGCTGCGGTGCAATCCGGA 1074
OY 1081 CTAGCCAAAGGCTTTTCGAGCAGCAGAGGGGCTTACTGTCGCAAGGGTGGCCACGAC 1140
Db 1075 CTAGCCAAAGGCTTTTCGAGCAGCAGAGGGGCTTACTGTCGCAAGGGTGGCCACGAC 1134
OY 1141 CTGGGGGTGCGGTTCACGAGCGCGCGGAGCGTCTGGCTGACTCTGCGCGCTGTTAC 1200
Db 1135 CTGGGGGTGCGGTTCACGAGCGCGCGGAGCGTCTGGCTGACTCTGCGCGCTGTTAC 1194
OY 1201 GCGGAGCAGCGTGGAGTAAAGGAAGGATGTCGGCG 1243
Db 1195 GCGGAGCAGCGTGGAGTAAAGGAAGGATGTCGGCG 1237

RESULT 10
LOCUS A89747 1237 bp DNA
DEFINITION Sequence 4 from Patent WO9832862.
ACCESSION A89747
VERSION A89747.1 GI:6738281
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1237)
AUTHORS Flohe, L. and Singh, M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
Location/Qualifiers
source 1..1237
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 394 c 386 g 221 t
ORIGIN

```

Query Match 97.98; Score 1219.4; E 5; Length 1237;  
 Best Local Similarity 99.48; Pred. No. 6.3e-108;

```

Matches 1236; Conservative 0; Mismatches 1; Indels 6; Gaps 1;
OY 1 ATCTTGAGATTAATGAACTTTCTTACACTGAGGCTACATATGAGAGGGTAATC 60
Db 1 ATCTTGAGATTAATGAACTTTCTTACACTGAGGCTACATATGAGAGGGTAATC 60
OY 61 ATGCGCTGCGTATTCGACGAGACCAAAAACAGCAATTCAGTTCCGGTGGCCATC 120
Db 61 ATGCGCTGCGTATTCGACGAGACCAAAAACAGCAATTCAGTTCCGGTGGCCATC 114
OY 121 ACCCGGCGCGGCTGCGGGAATTAACCGCTGTGGCCATGAGGTGCATCCAGGAGGT 180
Db 115 ACCCGGCGCGGCTGCGGGAATTAACCGCTGTGGCCATGAGGTGCATCCAGGAGGT 174
OY 181 GCGGAGAGAG 3CTGCGCTATCACCGAGCGGATTTCAAGGCGGAGCGGCAACTGTTC 240
Db 175 GCGGAGAGAGGCTCGGCTA. CACGAGCGGATTTCAAGGCGGAGCGGCAACTGTTC 234
OY 241 GCGACCGCGGACAGGTTGTGGCGGAGCGTGTATTTGCTCAAGGTCAAGAAGCGAT 300
Db 235 GCGACCGCGGACAGGTTGTGGCGGAGCGTGTATTTGCTCAAGGTCAAGAAGCGAT 294
OY 301 GCGGCGGAATACGGCGCGCTGCGACAGGCGAGATCTTGTTCAGTTCCTTCATTTGGCC 360
Db 295 GCGGCGGAATACGGCGCGCTGCGACAGGCGAGATCTTGTTCAGTTCCTTCATTTGGCC 354
OY 361 GCGTACGCTGCTTGACCGCATGCTGTGTGTGATTCGGGACACGCTAATGGCTACAG 420
Db 355 GCGTACGCTGCTTGACCGCATGCTGTGTGTGATTCGGGACACGCTAATGGCTACAG 414
OY 421 ACCGTCAGACCGCGGAGGCGGAGCTACCGCTGCTGCGCGATGAGCGAAGTGGCGGT 480
Db 415 ACCGTCAGACCGCGGAGGCGGAGCTACCGCTGCTGCGCGATGAGCGAAGTGGCGGT 474
OY 481 CGACTCGCGCGCCAGGTGGGCTTACCATGATGCGAAGGAGGGCGCGGTGTG 540
Db 475 CGACTCGCGCGCCAGGTGGGCTTACCATGATGCGAAGGAGGGCGCGGTGTG 534
OY 541 CTGATGGGCGGGGTGCGCGGGTGAACCGGCGAGCGTGTGTGATCGGGCGCGGAGC 600
Db 535 CTGATGGGCGGGGTGCGCGGGTGAACCGGCGAGCGTGTGTGATCGGGCGCGGAGC 594
OY 601 GCGGCTCAACAGCAGCGCTGATGCGCAACGAGTGGGCGGAGCGGCTTACGTTCTAG 660
Db 595 GCGGCTCAACAGCAGCGCTGATGCGCAACGAGTGGGCGGAGCGGCTTACGTTCTAG 654
OY 661 ATCAACATGACAACTTGTGCTGAGCTGAGCGCGGATCTGCGCGGATCCAACTGCG 720
Db 655 ATCAACATGACAACTTGTGCTGAGCTGAGCGCGGATCTGCGCGGATCCAACTGCG 714
OY 721 TACTCATGCGGCTACGAGCTGAGGGTCCGTCAAAGTTCGCACTGTATGGGGCC 780
Db 715 TACTCATGCGGCTACGAGCTGAGGGTCCGTCAAAGTTCGCACTGTATGGGGCC 774
OY 781 GTCTGTGTCGAGGGCGGCAAGGACCAAAATTAGTCTGAATTCCTTGTGGCCATATG 840
Db 775 GTCTGTGTCGAGGGCGGCAAGGACCAAAATTAGTCTGAATTCCTTGTGGCCATATG 834
OY 841 AAACGAGTGGGTACTGTGTATATAGCATAGCAGCAGGCGGGCTGTTCGAAGGCTCA 900
Db 835 AAACGAGTGGGTACTGTGTATATAGCATAGCAGCAGGCGGGCTGTTCGAAGGCTCA 894
OY 901 CGACGACCACTACGACCAACCGGAGCTGCGCGTGCAGAGACAGGTTTAAAGCGTG 960
Db 895 CGACGACCACTACGACCAACCGGAGCTGCGCGTGCAGAGACAGGTTTAAAGCGTG 954
OY 961 GCGAATGCGCGCTCGGTGCGGAGAGAGTGCAGCTACGCGCTGATACCAAGCGAGATG 1020
Db 955 GCGAATGCGCGCTCGGTGCGGAGAGAGTGCAGCTACGCGCTGATACCAAGCGAGATG 1014
OY 1021 CGGTATGTGCTGAGCTTGGCGACATGCTGCGGGGGGCGGCTGCGGTGCAATCCGGA 1080
Db 1015 CGGTATGTGCTGAGCTTGGCGACATGCTGCGGGGGGCGGCTGCGGTGCAATCCGGA 1074

```







421 ACCGTCAGACCGCCGAGCGGACGACCTACCCCTGCTGGCCCGGATGAGCGAAGTCCCGGT 480  
 415 ACCGTCAGACCGCCGAGCGGACGACCTACCCCTGCTGGCCCGGATGAGCGAAGTCCCGGT 474  
 481 CGACTCGCCCGCCAGGTTGGCGCTTACACCTGATGCGAACCCAGGGGGCCCGGTGTG 540  
 475 CGACTCGCCCGCCAGGTTGGCGCTTACACCTGATGCGAACCCAGGGGGCCCGGTGTG 534  
 541 CTGATGAGGCGGGGTGCGCCGCGCTCGAACCGGCGGAGGTGTGTGTCGCGCGCACAC 600  
 535 CTGATGAGGCGGGGTGCGCCGCGCTCGAACCGGCGGAGGTGTGTGTCGCGCGCACAC 594  
 601 GCGGCTACAGACGACCGCCGATCGCCAAAGGCGATGGCGCGGACCGTTACGCTTACAG 660  
 595 GCGGCTACAGACGACCGCCGATCGCCAAAGGCGATGGCGCGGACCGTTACGCTTACAG 654  
 661 ATACATGACAAATTTGGGCACTGACCGCGAGTCTGGGGCGCGGATCCACACTGCG 720  
 655 ATACATGACAAATTTGGGCACTGACCGCGAGTCTGGGGCGCGGATCCACACTGCG 714  
 721 TACTCATGCGCTACGAGCTCGAGGCTGCGTAAAGTCCGACCTGCTGATGGGCGC 780  
 715 TACTCATGCGCTACGAGCTCGAGGCTGCGTAAAGTCCGACCTGCTGATGGGCGC 774  
 781 GTCCTGCTGCGAGCGCGCAAGGCAACCAATTTAGTCTGCAATTCATTTGCGCGCATATG 840  
 775 GTCCTGCTGCGAGCGCGCAAGGCAACCAATTTAGTCTGCAATTCATTTGCGCGCATATG 834  
 841 AAACGAGTGGGCTACGAGTGGGATATAGCATGACGAGGCGGCTGTTTCGAAAGGCTCA 900  
 835 AAACGAGTGGGCTACGAGTGGGATATAGCATGACGAGGCGGCTGTTTCGAAAGGCTCA 894  
 901 CGACCGACGACCTACGACGACCGGAGCTTGGCGTGCAGC VCAAGCTGTTTACTGCGTG 960  
 895 CGACCGACGACCTACGACGACCGGAGCTTGGCGTGCAGC VCAAGCTGTTTACTGCGTG 954  
 961 GCGAATCGCCCGCTGCTGGTGGCGGAGAGCTTCGACCTGACGCGGCTGACCAAGCGAGATG 1020  
 955 GCGAATCGCCCGCTGCTGGTGGCGGAGAGCTTCGACCTGACGCGGCTGACCAAGCGAGATG 1014  
 1021 CCGATATGCTCGAGCTTGGCGACGACGATGCTGGGCGGCGCTGCGGCTGCAATCCGCGCA 1080  
 1015 CCGATATGCTCGAGCTTGGCGACGACGATGCTGGGCGGCGCTGCGGCTGCAATCCGCGCA 1074  
 1081 CTACCCAAAGGTCTTTCGAGCGACGAGAGGCGCTTACTGTCGAAAGGCTGGCGAC 1140  
 1075 CTACCCAAAGGTCTTTCGAGCGACGAGAGGCGCTTACTGTCGAAAGGCTGGCGAC 1134  
 1141 CTGGGGGTGCGCTTACCGAGCGCGGAGGCTGCTGGCTGACCTGCGCGCTGTTAC 1200  
 1135 CTGGGGGTGCGCTTACCGAGCGCGGAGGCTGCTGGCTGACCTGCGCGCTGTTAC 1194  
 1201 GCGGAGCAGACGTCGAGGATTAAGGAGGCGATGA 1234  
 1195 GCGGAGCAGACGTCGAGGATTAAGGAGGCGATGA 1228

RESULT 14  
 LOCUS A89748 1228 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 5 from Patent WO9832862.  
 ACCSSION A89748  
 VERSION A89748.1 GI:6738282  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 unclassified.

REFERENCE  
 1 (bases 1 to 1228)  
 Fiohe L. and Singh M.  
 L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
 Patent: WO 9832862-A 30-JUL-1998.  
 FIOHE LEOPOLD (DE); SINGH MAHAVIR (DE)

FEATURES  
 source location/Qualifiers  
 1..1228 /organism="unidentified"  
 /db\_xref="taxon:3264"  
 BASE COUNT 236 a 391 c 382 g 219 t  
 ORIGIN

Query Match 97.2%; Score 1210.4; DB 5; Length 1228;  
 Best Local Similarity 99.4%; Pred. No. 1.3e-156;  
 Matches 1227; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

1 ATCTGAGATTAATCGAATCTTCTTCACTAGAGCTGACGATGAGAGGGGTAATC 60  
 1 ATCTGAGATTAATCGAATCTTCTTCACTAGAGCTGACGATGAGAGGGGTAATC 60  
 61 ATGCGGCTGCGTATTTCCG ACCGAGACCAAAACAAAGCAATTCATTCGCGGTGCCATC 120  
 61 ATGCGGCTGCGTATTTCCG ACCGAGACCAAAACAAAGCAATTCATTCGCGGTGCCATC 114  
 121 ACCCGGCGCGGCTGCGGGAATTAACCGCTGCGCATGAGTGGCTCATCCAGGAGGT 180  
 115 ACCCGGCGCGGCTGCGGGAATTAACCGCTGCGCATGAGTGGCTCATCCAGGAGGT 174  
 181 GCGGAGAGGCGCTGCGGCTATCACCAGCGGATTTCAAGGCGGAGCGGCAACTGCTC 240  
 175 GCGGAGAGGCGCTGCGGCTATCACCAGCGGATTTCAAGGCGGAGCGGCAACTGCTC 234  
 241 GGCACCGCGGACGAGTGTGGCGGAGCGTGTATTTGCTCAAGGTCAAGAAACGATA 300  
 235 GGCACCGCGGACGAGTGTGGCGGAGCGGATTTATTTGCTCAAGGTCAAGAAACGATA 294  
 301 GCGGAGAGTACGCGGCGCTGCGGAGCGGAGATCTTTCAGCTTTCATTTGGCC 360  
 295 GCGGAGAGTACGCGGCGCTGCGGAGCGGAGATCTTTCAGCTTTCATTTGGCC 354  
 355 GCGTACGCGTGTGACCGGATGCGGATTTGGATTCGCGACACGTCATATGCTACGAG 414  
 361 GCGTACGCGTGTGACCGGATGCGGATTTGGATTCGCGACACGTCATATGCTACGAG 420  
 421 ACCGTCAGACCGCGGAGCGGACGACCTGCTGCTGGCCGAGTGAAGGAGTGGCGGT 480  
 415 ACCGTCAGACCGCGGAGCGGAGCGGACGACCTGCTGCTGGCCGAGTGAAGGAGTGGCGGT 474  
 481 CGACTCGCGCGCCAGGTTGCGCTTACCACTGATGCAACCAAGGCGGCGGCTGTG 540  
 475 CGACTCGCGCGCCAGGTTGCGCTTACCACTGATGCAACCAAGGCGGCGGCTGTG 534  
 541 CTGATGAGGCGGCGTCCGCGCTGCAACCGCGGAGCGTGTGATGCGCGCGCAC 600  
 535 CTGATGAGGCGGCGTCCGCGCTGCAACCGCGGAGCGTGTGATGCGCGCGCAC 594  
 601 GCGGCTACAGACGACCGCGGATGCGCAAGGCGATGGGCGGAGCGTACGATTTAGAC 660  
 595 GCGGCTACAGACGACCGCGGATGCGCAAGGCGATGGGCGGAGCGTACGATTTAGAC 654  
 661 ATCAACATGACAAATTTGCGCACTGCAAGCGCGGATTTGCGGCGGAGTCAACTGCG 720  
 655 ATCAACATGACAAATTTGCGCACTGCAAGCGCGGATTTGCGGCGGAGTCAACTGCG 714  
 721 TACTCATGCGCTACGAGCTCGAGGCTGCGTCAACAGTGGCGAGCTGATTTGGGCGC 780  
 715 TACTCATGCGCTACGAGCTCGAGGCTGCGTCAACAGTGGCGAGCTGATTTGGGCGC 774  
 781 GTCCTGCTGCGAGCGCGCAAGGCAACCAATTTAGTCTGCAATTCATTTGCGCGCATATG 840  
 775 GTCCTGCTGCGAGCGCGCAAGGCAACCAATTTAGTCTGCAATTCATTTGCGCGCATATG 834  
 841 AAACGAGTGGGCTACTGATGATATAGCATGACGAGCGGCTGTTTGAAGGCTCA 900  
 835 AAACGAGTGGGCTACTGATGATATAGCATGACGAGCGGCTGTTTGAAGGCTCA 894  
 901 CGACGACGACCTACGACGACCGGAGCTTGGCGTGCAGACGACGCTTTTACTGCGTG 960

```

Db 895 GAGCGACACTAGACACACCGGCTGCGCGGACGACAGCCTGTTACTGCGTG 954
Qy 961 GCGAATATGCGCGCTCGCTGCGGAGAGAGCTGACCTGACGCGGACGACGAGT 1020
Db 955 GCGAATATGCGCGCTCGCTGCGGAGAGAGCTGACCTGACGCGGACGAGT 1014
Qy 1021 CCGATGCTGCTGAGCTGCGGACGACGCTGCGGAGAGAGCTGACCTGACGAG 1080
Db 1015 CCGATGCTGCTGAGCTGCGGACGACGCTGCGGAGAGAGCTGACCTGACGAG 1074
Qy 1081 CTAGCCAAAGGCTCTTTCAGACGACGAGAGGCGCTGCTGCTGACGAGGCTGAC 1140
Db 1075 CTAGCCAAAGGCTCTTTCAGACGACGAGAGGCGCTGCTGCTGACGAGGCTGAC 1134
Qy 1141 CTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db 1135 CTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1194
Qy 1201 GCGGAGCAGCAGCTGCGGAGTAAGGAGGAGTGA 1234
Db 1195 GCGGAGCAGCAGCTGCGGAGTAAGGAGGAGTGA 1228

RESULT 15
A87609 1235 bp DNA PAT 22-JAN-2000
LOCUS A87609 Sequence 6 from Patent WO9836089.
DEFINITION A87609
ACCESSION A87609
VERSION A87609.1 GI:6736249
KEYWORDS
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1235)
AUTHORS Fiohe, L. and Singh, M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FIOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
Source Location/Qualifiers
1..1235
/organism="unclassified"
/db_xref="taxon:32644"
BASE COUNT 235 a 395 c 384 g 220 t 1 others
ORIGIN
Query Match 96.9%; Score 1207; DB 5; Length 1235;
Best Local Similarity 99.4%; Pred. No. 3; 9e-166;
Matches 1234; Conservative 0; Mismatches 1; Indels 7; Gaps 2;

```

```

Db 295 GCGGCGAATACGCGCGCTGCGACAGGGC -GATCTTGTCACTTCTTGGATTTGGCC 353
Qy 361 GCGTACAGCTCTTGCACACGATGCTGTTGATTTCCGGACACAGTGAATTTGCTACGAG 420
Db 354 GCGTACAGCTCTTGCACACGATGCTGTTGATTTCCGGACACAGTGAATTTGCTACGAG 413
Qy 421 ACCGTCCAGACCGCGGACGCGGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
Db 414 ACCGTCCAGACCGCGGACGCGGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 473
Qy 481 CGACTGCGCGCCAGGTTGGGCTTACCTGATGCGAAGCC -AAGGAGGCGCGGCTGCTG 540
Db 474 CGACTGCGCGCCAGGTTGGGCTTACCTGATGCGAAGCC -AAGGAGGCGCGGCTGCTG 533
Qy 541 CTGATGCGCGGCTGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 600
Db 534 CTGATGCGCGGCTGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 593
Qy 601 GCGGCTTACAAAGCGAGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 660
Db 594 GCGGCTTACAAAGCGAGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 653
Qy 661 ATCAACATCGACAACTTTCGCAACTGAGCGAGTTCGCGGCGGAGTTCACACTGCG 720
Db 654 ATCAACATCGACAACTTTCGCAACTGAGCGAGTTCGCGGCGGAGTTCACACTGCG 713
Qy 721 TACTCATGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 780
Db 714 TACTCATGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 773
Qy 781 GTCTGTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 840
Db 774 GTCTGTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 833
Qy 841 AAACGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Db 834 AAACGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 893
Qy 901 GCGGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 894 GCGGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 953
Qy 961 GCGAATATGCGCGCTCGCTGCGGAGAGAGCTGACCTGACGCTGACGACGAGGAGT 1020
Db 954 GCGAATATGCGCGCTCGCTGCGGAGAGAGCTGACCTGACGCTGACGACGAGGAGT 1013
Qy 1021 CCGATGCTGCTGAGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 1080
Db 1014 CCGATGCTGCTGAGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 1073
Qy 1081 CTAGCCAAAGGCTCTTTCAGACGACGAGAGGCGCTTACTGCTGCGAAGGCTGCGGCTGCGGCTGCG 1140
Db 1074 CTAGCCAAAGGCTCTTTCAGACGACGAGAGGCGCTTACTGCTGCGAAGGCTGCGGCTGCGGCTGCG 1133
Qy 1141 CTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db 1134 CTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1193
Qy 1201 GCGGAGCAGCAGCTGCGGAGTAAGGAGGAGTGTGCGGC 1242
Db 1194 GCGGAGCAGCAGCTGCGGAGTAAGGAGGAGTGTGCGGC 1235

```

Search completed: June 22, 2000, 14:50:58  
Job time: 17473 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:04:40 ; Search time 458.59 Seconds

(without alignments)  
679.232 Million cell updates/sec

Title: US-09-362-485-2

Sequence: 1 ATCTTCGACATTAATCGAAC.....AAGCATGATGCGCGCG 1245

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result bel., printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1245	100.0	1245	1 V49626	Mycobacterium tube
2	1245	100.0	1260	1 V49510	Mycobacterium sp.
3	1245	100.0	1260	1 V49625	Mycobacterium tube
4	467.6	37.6	682	1 V49511	Mycobacterium mari
5	261.2	21.0	1125	1 N91423	Sequence of heat-r
6	198.4	15.9	28171	1 V52155	Streptococcus pneu
7	178.4	14.3	9280	1 V74442	Streptococcus aur
8	94.8	7.6	1074	1 T67971	H. pylori membrane
9	69.2	5.6	544	1 V43039	Streptococcus pneu
10	57.4	4.6	31096	1 V74370	Staphylococcus aur
11	50.8	4.1	390	1 021833	Randomising oligon
12	50.8	4.1	390	1 036859	PCR primer for 5'
13	50.4	4.0	15872	1 T68715	Streptomycetes venez
14	49	3.9	1269	1 003065	Sequence encoding
15	47.6	3.8	1413	1 X34205	Mycobacterium spec
16	47.6	3.8	1722	1 X34206	Mycobacterium spec
17	47.6	3.8	3946	1 T93610	Mycobacterium tube
18	47.6	3.8	28598	1 T06769	Sorangium cellulos
19	47.6	3.8	28958	1 T89955	Sorangium cellulos
20	47.6	3.8	49377	1 T05287	The soraphen biosy
21	46.4	3.7	985	1 V44439	Mycobacterium tube
22	46.4	3.7	985	1 V64548	M. tuberculosis im
23	46	3.7	1520	1 Q22482	gREL-1 gene codin
24	46	3.7	1520	1 T73117	Acetabularia sp. A
25	46	3.7	2668	1 Q22485	gREL-1 gene. Reco
26	46	3.7	17955	1 V56642	Actinoplanes sp. a
27	45.4	3.6	833	1 064203	snab gene encoding
28	45.4	3.6	5392	1 064201	Sequence compisin
29	42.4	3.4	1833	1 064206	snR gene encoding
30	42.4	3.4	2185	1 V84066	Clone p5-4 encodin
31	42.4	3.4	2186	1 093655	S. lividans protea
32	42.4	3.4	24379	1 T93095	Streptomyces fireo
33	42.4	3.4	24379	1 V25925	Streptomyces roseo
34	42.2	3.4	329	1 V44425	Mycobacterium tube

35	42.2	3.4	329	1 V64534	M. tuberculosis im
36	42	3.4	3946	1 T93610	Mycobacterium tube
37	42	3.4	4018	1 063879	Polydactyloalkano
38	41.8	3.4	882	1 V44403	Mycobacterium tube
39	41.8	3.4	882	1 V64512	M. tuberculosis im
40	41.6	3.3	12036	1 004658	FHA structural gen
41	41.2	3.3	2414	1 005926	Sequence encoding
42	41	3.3	29879	1 046806	eryA region of S.
43	40.8	3.3	2151	1 076252	HSV-2 protease, IC
44	40.8	3.3	2151	1 076261	HSV-2 protease/ICP
45	40.8	3.3	2472	1 084671	HSV-2 UL26 gene. N

## ALIGNMENTS

RESULT 1	
V49626	standard; DNA; 1245 BP.
AC V49626:	
DE 20-NOV-1998 (first entry)	
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.	
KW ss: Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.	
OS Mycobacterium tuberculosis.	
PN W09836089-A2.	
PD 20-AUG-1998.	
PF 20-JAN-1998; E00483.	
PR 29-JAN-1997; EP-101338.	
PA (FLOH/) FLOHE L.	
PI Flohe L, Hutter B, Kolz A, Singh M;	
DR WPI: 98-457123/39.	
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity	
PT - useful for, e.g. for diagnosis, differentiation of strains,	
PT monitoring vaccination and identification of mycobacterial	
PT inhibitors	
PS Disclosure: Fig 3.19; 55pp; German.	
CC The Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in	
CC the production of kits for diagnosing tuberculosis (TB) and other	
CC mycobacterial infections in humans or animals. Kits are used for direct	
CC diagnosis of TB on clinical samples (e.g. body fluids) and can	
CC differentiate between pathogenic and non-virulent strains, e.g. for	
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may	
CC also be used to identify substances that inhibit mycobacteria, for	
CC combatting epidemics and for vaccination follow-up. Oligonucleotides	
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,	
CC also for culture confirmation of isolated strains and for chromosome	
CC fingerprinting to detect/differentiate between mycobacteria, and for	
CC L-alanine-specific biotransformation reactions. AlaDH is an early	
CC antigen, present extracellularly after only a few days of growth, making	
CC it an ideal drug target.	
SO Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;	

Query Match 100.0%; Score 1245; DB 1; Length 1245;  
Best Local Similarity 100.0%; Pred. No. 2.3e-251;  
Matches 1245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1	ATCTTCGACATTAATCGAATCTTCTTCACTGAAGCGTACGATGAGAGGGGTATC 60
DB 1	ATCTTCGACATTAATCGAATCTTCTTCACTGAAGCGTACGATGAGAGGGGTATC 60
QY 61	ATGGCGGCGGTATTCGACGAGCAACCAAAACGAATTCGAATTCGGGTGGCATC 120
DB 61	ATGGCGGCGGTATTCGACGAGCAACCAAAACGAATTCGAATTCGGGTGGCATC 120
QY 121	ACCCGCGCGGCGGTGCGGAACTAACCGCGTGTGGCCATGAGTGTCTATCGAGCGGT 180
DB 121	ACCCGCGCGGCGGTGCGGAACTAACCGCGTGTGGCCATGAGTGTCTATCGAGCGGT 180
QY 181	GCCGAGAGGGCGTGGGATACCGAGCGGATTTCAAGGGGAGGCGGCAACTGTC 240
DB 181	GCCGAGAGGGCGTGGGATACCGAGCGGATTTCAAGGGGAGGCGGCAACTGTC 240

241 GGCACCGCCGACAGTGTGGGCGGACGCTGATTATTTGCTCAAGGTCMAAGAACGATA 300  
 241 GGCACCGCCGACAGTGTGGGCGGACGCTGATTATTTGCTCAAGGTCMAAGAACGATA 300  
 301 GCGGCGGAATACGGCCGCTCGACACGGGAGATCTTGTTCAGTCTTCCATTGGCC 360  
 301 GCGGCGGAATACGGCCGCTCGACACGGGAGATCTTGTTCAGTCTTCCATTGGCC 360  
 301 GCGGCGGAATACGGCCGCTCGACACGGGAGATCTTGTTCAGTCTTCCATTGGCC 360  
 361 GCGTACGCTGTGACGAGTGTGTTGATTCCGACACGATGATGCTACGATGCTAGAG 420  
 361 GCGTACGCTGTGACGAGTGTGTTGATTCCGACACGATGATGCTACGATGCTAGAG 420  
 421 ACCGTCCAGACCGCCGACGAGGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
 421 ACCGTCCAGACCGCCGACGAGGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
 481 GAGCTCGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 481 GAGCTCGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 541 CTGATGGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
 541 CTGATGGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
 601 GCGGCGTCAACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
 601 GCGGCGTCAACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
 661 ATCAACATGACAACTTCCGCACTGCAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
 661 ATCAACATGACAACTTCCGCACTGCAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
 721 TACTATGCGGCTTACGAGTGTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
 721 TACTATGCGGCTTACGAGTGTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
 781 GTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
 781 GTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
 841 AAACGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
 841 AAACGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
 901 GCGGCGTCAACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
 901 GCGGCGTCAACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
 961 GCGGCGTCAACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
 961 GCGGCGTCAACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
 1021 GCGGCGTCAACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080  
 1021 GCGGCGTCAACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080  
 1081 CTAGCCAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140  
 1081 CTAGCCAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140  
 1141 CTGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200  
 1141 CTGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200  
 1201 GCGGCGTCAACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
 1201 GCGGCGTCAACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260

20-OCT-1998 (first entry)  
 Mycobacterium sp. AladH DNA.  
 AladH: AladH; diagnosis: tuberculosis; pathogen;  
 swimmers disease; vaccine; epidemic; infection; identification; ss.  
 Mycobacterium sp.  
 20-JUL-1998.  
 29-JAN-1998; E00484.  
 29-JAN-1997; EP-101339.  
 PI (FLOH) FLOHE L.  
 PI Flohe L. Hutter B. Kolik A. Singh M.  
 DR WPI: 98-427958/36.  
 PT Nucleic acid encoding alanine dehydrogenase (ADH) protein isolated  
 PT - used for diagnosis of tuberculosis and other mycobacterial  
 PT diseases, also for treatment and prevention, for drug screening and  
 PT for bio-transformation  
 PS Disclosure: Page 11: 57pp; German.  
 CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated  
 CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis  
 CC and other mycobacterial infections (including 'swimmers' disease', caused  
 CC by M. marinum, a fish pathogen) in humans or animals. The protein can  
 CC also be used for control of epidemics and for vaccination, to screen for  
 CC agents with anti-mycobacterial activity, and in bio-transformations that  
 CC are specific for L-alanine. Also mycobacteria can be identified by  
 CC analysis of genomic ADH sequences. ADH is an antigen that is secreted  
 CC early during infection. 243 A: 403 C: 389 G: 225 T:  
 Sequence 1260 BP: 243 A: 403 C: 389 G: 225 T:

Query Match 100.0%; Score 1245; DB 1; Length 1260;  
 Best Local Similarity 100.0%; Pred. No. 2,3e-251;  
 Matches 1245; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

1 ATCTTGAGATTAATGCACTTCTTACACTGAGGCTACATGAGGAGGATTC 60  
 16 ATCTTGAGATTAATGCACTTCTTACACTGAGGCTACATGAGGAGGATTC 75  
 61 ATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
 76 ATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 135  
 121 ACCCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180  
 136 ACCCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 195  
 181 GCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240  
 196 GCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 255  
 241 GGCACCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300  
 256 GGCACCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 315  
 301 GCGGCGGATACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
 316 GCGGCGGATACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 375  
 376 GCGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 435  
 421 ACCGTCCAGACCGCCGACGAGGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
 436 ACCGTCCAGACCGCCGACGAGGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 495  
 481 GAGCTCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 496 GAGCTCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 555  
 541 CTGATGGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
 556 CTGATGGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 615

RESULT 2  
 V49510  
 ID V49510 standard; DNA; 1260 BP.  
 AC V49510;



QY 601 GCCGCTTAACGACGACCCCGCATCCGCAACGGATGGGGCGACCGTCTACGGTTCTAGAC 660  
 DB 616 GCGGCTTACAGCAGCAGCCGCGATCCGCAACGGATGGGGCGACCGTCTACGGTTCTAGAC 675  
 QY 661 ATCAACATGACAAATTTGGGCACTGAGCGCGATGTTGCGGGCGGATTCACATCCG 720  
 DB 676 ATCAACATGACAAATTTGGGCACTGAGCGCGATGTTGCGGGCGGATTCACATCCG 735  
 QY 721 TACTATCGGCTTACGAGCTCGAGGGTGGCCGTAACGTCGCGACTGTGATGGGGCC 780  
 DB 736 TACTATCGGCTTACGAGCTCGAGGGTGGCCGTAACGTCGCGACTGTGATGGGGCC 795  
 QY 781 GTCTGTGTCGAGCGCGCAAGGCAATTTGTCTCGAATTCACCTTGTGCGGGCATATG 840  
 DB 796 GTCTGTGTCGAGCGCGCAAGGCAATTTGTCTCGAATTCACCTTGTGCGGGCATATG 855  
 QY 841 AAACGAGTGGCTACTGGTGGATATAGCATGACGAGGCGGTGTTTCGAAGCTCA 900  
 DB 856 AAACGAGTGGCTACTGGTGGATATAGCATGACGAGGCGGTGTTTCGAAGCTCA 915  
 QY 901 CGACGACGACCTTACGACGACCGGAGTTGCGCGTGCACAGACGCTGTTTACTGCGTG 960  
 DB 916 CGACGACGACCTTACGACGACCGGAGTTGCGCGTGCACAGACGCTGTTTACTGCGTG 975  
 QY 961 GCGACATGCGCGCTGCTGGTGGCGAAGAGCTGACCTAGCGCTGACCAACGGGAGCATG 1020  
 DB 976 GCGACATGCGCGCTGCTGGTGGCGAAGAGCTGACCTAGCGCTGACCAACGGGAGCATG 1035  
 QY 1021 CGCTATGCTGCTGAGCTGGCGACGATGGTGGGCGCGCTGCGGCTGATTCGCGCA 1080  
 DB 1036 CGCTATGCTGCTGAGCTGGCGACGATGGTGGGCGCGCTGCGGCTGATTCGCGCA 1095  
 QY 1081 CTAGCCAAAGTCTTTCGACGACGAGAGGGCGTCTACTGTCCGAACGGTGGCCACCGAC 1140  
 DB 1096 CTAGCCAAAGTCTTTCGACGACGAGAGGGCGTCTACTGTCCGAACGGTGGCCACCGAC 1155  
 QY 1141 CTGGGGGTCGCTTTCACCGAGCGCGGAGCTGCTGGCTGACTCTGCGCGCTCTTAC 1200  
 DB 1156 CTGGGGGTCGCTTTCACCGAGCGCGGAGCTGCTGGCTGACTCTGCGCGCTCTTAC 1215  
 QY 1201 GCGGAGCAGCTGGGAGTAAAGGAAGCATGTGTCGCGCGG 1245  
 DB 1216 GCGGAGCAGCTGGGAGTAAAGGAAGCATGTGTCGCGCGG 1260  
 RESULT 3  
 V49625  
 ID V49625 standard: DNA: 1260 BP.  
 AC V49625;  
 DT 20-NOV-1998 (first entry)  
 DE Mycobacterium tuberculosis: Alanine dehydrogenase.  
 KW ss: Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.  
 OS Mycobacterium tuberculosis.  
 PN M09836088-A2.  
 PD 20-AUG-1998.  
 PR 29-JAN-1997; E00483.  
 PA (FLOH) FLOHE L.  
 PI Flohe L, Hutter B, Kolk A, Singh M;  
 DR WPI: 98-457123/39.  
 PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity  
 PT useful for, e.g. for diagnosis differentiation of strains,  
 PT monitoring vaccination and identification of mycobacterial  
 PT inhibitors  
 CC Claim 13; Fig 2.3; 55pp; German.  
 CC The Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in  
 CC the production of kits for diagnosing tuberculosis (TB) and other  
 CC mycobacterial infections in humans or animals. Kits are used for direct  
 CC diagnosis of TB on clinical samples (e.g. body fluids) and can  
 CC differentiate between pathogenic and non-virulent strains, e.g. for  
 CC identifying pathogens of the M. tuberculosis (M. t.) complex. The kit may  
 CC also be used to identify substances that inhibit mycobacteria, for  
 CC combatting epidemics and for vaccination follow-up. Oligonucleotides

CC derived from AlaDH are used similarly in diagnostic hybridisation tests,  
 CC also for culture confirmation of isolated strains and for chromosome  
 CC fingerprinting to detect/differentiate between mycobacteria, and for  
 CC L-alanine-specific biotransformation reactions. AlaDH is an early  
 CC antigen, present extracellularly after only a few days of growth, making  
 CC it an ideal drug target.  
 SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;  
 Query Match 100.0%; Sec: 12.5; DB 1; Length 1260;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-251;  
 Matches 1245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACTTTCACATTAATTCAGACTTCTTTCACACTGAGACCGTACAGATTCAGAGGGTATTC 60  
 DB 16 ATCTTTCACATTAATTCAGACTTCTTTCACACTGAGACCGTACAGATTCAGAGGGTATTC 75  
 QY 61 ATGCGGCTGGTATTCCGACCGGACCAAAACAGATTCATTCGCGGTGGCCATC 120  
 DB 76 ATGCGGCTGGTATTCCGACCGGACCAAAACAGATTCATTCGCGGTGGCCATC 135  
 QY 121 ACCCGGCGCGGCTCGGGAAGTAAACCGTGTGCGCATGAGTGTCTATCCAGCGAGT 180  
 DB 136 ACCCGGCGCGGCTCGGGAAGTAAACCGTGTGCGCATGAGTGTCTATCCAGCGAGT 195  
 QY 181 GCGGAGAGGCTCGGCTATACCGGACGCGGATTTCAAGCGCGGAGCGCGGACACTGTC 240  
 DB 196 GCGGAGAGGCTCGGCTATACCGGACGCGGATTTCAAGCGCGGAGCGCGGACACTGTC 255  
 QY 241 GGCACCGCGGACAGGTGTGGGCGGACGCTGATTTATTCCTCAAGGTCAAAAGCGATA 300  
 DB 256 GGCACCGCGGACAGGTGTGGGCGGACGCTGATTTATTCCTCAAGGTCAAAAGCGATA 315  
 QY 301 GCGGCGGAATACGCGCGCTCGGACGCGGAGATTTGTACGTTCTTGCATTTGGCC 360  
 DB 316 GCGGCGGAATACGCGCGCTCGGACGCGGAGATTTGTACGTTCTTGCATTTGGCC 375  
 QY 361 GCGTCACTGCTTTCACCGCATGCTGTTGGATTCGCGGACGACGTCATTCGATCGAG 420  
 DB 376 GCGTCACTGCTTTCACCGCATGCTGTTGGATTCGCGGACGACGTCATTCGATCGAG 435  
 QY 421 ACCGTCCAGACGCGGACGCG :CACTACCGCTGCTGCGCGCATGAGCGAAGTCGCGGT 480  
 DB 436 ACCGTCCAGACGCGGACGCGGACGCTACCTGCTTGTCCCATGAGCGAAGTCGCGGT 495  
 QY 481 CGACTCGCGCGCGGCTTTCACCGCATGCTGAGCAACCGAAGGGGCGCGGTGTG 540  
 DB 496 CGACTCGCGCGCGGCTTTCACCGCATGCTGAGCAACCGAAGGGGCGCGGTGTG 555  
 QY 541 CTGATGGGCGGCTGCTTCGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
 DB 556 CTGATGGGCGGCTGCTTCGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 615  
 QY 601 GCGGCTTACAGCGAGCGCGCATGCGCAACGGATGGGCGCGCATGCGGTTCTAGAC 660  
 DB 616 GCGGCTTACAGCGAGCGCGCATGCGCAACGGATGGGCGCGCATGCGGTTCTAGAC 675  
 QY 661 ATCAACATGACAAATTTGGGCACTGAGCGCGATGTTGCGGGCGGATTCACATCCG 720  
 DB 676 ATCAACATGACAAATTTGGGCACTGAGCGCGATGTTGCGGGCGGATTCACATCCG 735  
 QY 721 TACTATCGGCTTACGAGCTCGAGGGTGGCCGTAACGTCGCGACTGTGATGGGGCC 780  
 DB 736 TACTATCGGCTTACGAGCTCGAGGGTGGCCGTAACGTCGCGACTGTGATGGGGCC 795  
 QY 781 GTCTGTGTCGAGCGCGCAAGGCAATTTGTCTCGAATTCACCTTGTGCGGGCATATG 840  
 DB 796 GTCTGTGTCGAGCGCGCAAGGCAATTTGTCTCGAATTCACCTTGTGCGGGCATATG 855  
 QY 841 AAACGAGTGGCTACTGGTGGATATAGCATGACGAGGCGGTGTTTCGAAGCTCA 900  
 DB 856 AAACGAGTGGCTACTGGTGGATATAGCATGACGAGGCGGTGTTTCGAAGCTCA 915



```

QY 901 GCACGACCACTAGACACACCGACCTTCCGCTGACAGACGCTGTTTACTGCTG 960
DB 916 CCACGACCACTAGACACACCGACCTTCCGCTGACAGACGCTGTTTACTGCTG 975
QY 961 GGAACATCCCGGCTGCTGCGAAGAGTGTGACTAGCGCTGACCAACCGAGATG 1020
DB 976 GGAACATCCCGGCTGCTGCGAAGAGTGTGACTAGCGCTGACCAACCGAGATG 1035
QY 1021 CCGTATGCTGAGCTTCCGACATGCTGCGGCGGCTGCGCTGCAATCCGCA 1080
DB 1036 CCGTATGCTGAGCTTCCGACATGCTGCGGCGGCTGCGCTGCAATCCGCA 1095
QY 1081 CTAAGCAAAAGCTTTTCCAGCAGCAGAGGGGCTTACTCTCCGAAGGGTGGCCACAC 1140
DB 1096 CTAAGCAAAAGCTTTTCCAGCAGCAGAGGGGCTTACTCTCCGAAGGGTGGCCACAC 1155
QY 1141 CTGGGGGCTGCTTCCAGCAGCAGAGGGGCTTACTCTCCGAAGGGTGGCCACAC 1200
DB 1156 CTGGGGGCTGCTTCCAGCAGCAGAGGGGCTTACTCTCCGAAGGGTGGCCACAC 1215
QY 1201 GCGGACACACGTCGGAGTAAGGAGCGATGATGCTGGCGC 1245
DB 1216 GCGGACACACGTCGGAGTAAGGAGCGATGATGCTGGCGC 1260

```

```

RESULT 4
V49511
ID V49511 standard; DNA: 682 BP.
AC V49511;
DE 20-OCT-1998 (first entry)
DE Mycobacterium marinum Mar3 DNA.
KM Alamine dehydrogenase; AlADH; ADH; diagnosis: tuberculosis; pathogen;
KM Swimmers disease; vaccine; epidemic; infection; identification: ss.
OS Mycobacterium marinum.
FH Key Location/Qualifiers
FT CDS 1..682
FT /tag= a
FT /codon_start= 3
FT /product= "Mar3"
FT /note= "Alamine dehydrogenase"
PN W09832862-A2.
PD 30-JUL-1998.
PE 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Fione L, Hutter B, Kolx A, Singh M;
DR WPI: 98-427958/36.
DR P-PSDB: W64481.
PT Nucleic acid encoding alamine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for +ug screening and
PT for bio-transformation
PS Claim 1; Page 34-35; 57pp; German.
CC This sequence encodes an alamine dehydrogenase (ADH) protein, Mar3
CC isolated from Mycobacterium marinum. This protein is used to diagnose
CC tuberculosis and other mycobacterial infections (including 'swimmers'
CC disease', caused by M. marinum, a fish pathogen) in humans or animals.
CC The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 682 BP; 105 A; 254 C; 225 G; 98 T;

```

```

Query Match 37.6%; Score 467.6; DB 1; Length 682;
Best Local Similarity 80.4%; Pred. No. 2,2e-89;
Matches 548; Conservative 0; Mismatches 134; Indels 0; Gaps 0;
QY 104 AATTCCGGTGGCTACACCCCGCGGCTGCGGAGTAACCCGCTGCTGCTGAG 163
DB 1 AATTCCGGTGGCTACACCCCGCGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTG 60

```

```

QY 164 TGCATCATCAGCAGAGTCCCGAGAGAGGCTCGGCTATCACCAGCGGATTTCAAGCG 223
DB 61 TGCATCATCAGCAGAGTCCCGAGAGAGGCTCGGCTATCACCAGCGGATTTCAAGCG 120
QY 224 CAGGCGCCCACTGTGCGCACCCGCCACAGGTGTGGCCGCTGATTTATGCTCA 283
DB 121 CCGGTGCCCACTGTATCAGACCCGCCACAGGTGTGGCCGCTGATTTATGCTCA 180
QY 284 AGGTAAAGAACCGATACGCGCGAATACGCGCGCTCGACAGCGGAGATCTTTTCA 343
DB 181 AGGTAAAGAACCGATACGCGCGAATACGCGCGCTCGACAGCGGAGATCTTTTCA 240
QY 344 GCTTCTCATTTTGGCGCGTACGCTGCTTGCACCGATCGTTTGTGATTCGCGACA 403
DB 241 CTTACTCCTACCTGCGCGCTGCGCGCTGCGCGCTGCGCGATGCTGTAAGTCCGCGACA 300
QY 404 GCTCATTTGCTACAGAGACCGTCCAGACCGCGGAGCGGAGCTACCTGCTTCCCGCA 463
DB 301 GCTCATTTGCTACAGAGACCGTCCAGACCGCGGAGCGGAGCTACCTGCTTCCCGCA 360
QY 464 TGAGCGAAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 523
DB 361 TGAGCGAAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
QY 524 AAGGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 583
DB 421 ACGTCGCTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
QY 584 TGATGCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 643
DB 481 TGATGCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
QY 644 CCGTACGCTTCTAGACATCAACATTCGCAACTTCGCAACTTCGCAACTTCGCA 703
DB 541 TGTACACGCTGCTGATGATCAACATCAACATTCGCAACTTCGCAACTTCGCA 600
QY 704 GCGGATCCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 763
DB 601 GTCGCTCGGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
QY 764 ACCTGCTGATGGGCGCTGCT 785
DB 661 ACATGCTGATGGGCGCTGCT 682

```

```

RESULT 5
N91423
ID N91423 standard; DNA: 1125 BP.
AC N91423;
DE 01-FEB-1991 (first entry)
DE Sequence of heat-resistant alamine dehydrogenase (AH) gene with mol. wt.
DE 2 MD or less
DE Enzyme: ds.
OS Bacillus stearothermophilus IFO 12550.
PN J01043194-A.
PD 15-FEB-1989.
PE 10-AUG-1987; 200524.
PR 10-AUG-1987; JP-200524.
PA (NIRA) Unitika KK.
DR WPI: 89-096096/13.
PT Recombinant plasmid for transforming Escherichia coli -
PT obd. by connecting heat resistant alamine dehydrogenase gene to
PT vector plasmid
PS Disclosure: Fig 2, p693; 9pp; Japanese.
CC A recombinant plasmid conf. heat-resistant alamine dehydrogenase (AH)
CC gene with mol. wt. 2 or less MD is claimed. Cells transformed with the
CC vector produce high levels of heat-resistant AH. Transformed E. coli
CC is useful as a clinical inspection reagent.
SQ Sequence 1125 BP; 246 A; 295 C; 367 G; 217 T;

```

```

Query Match 21.0%; Score 261.2; DB 1; Length 1125;
Best Local Similarity 55.9%; Pred. No. 2,4e-46;

```

	Matches	604:	Conservative	0:	Mismatches	458:	Indels	18:	Gaps
QY	61	ATGCGCGTGTGATTATCCAGACCGAGACCAAAAACAACGAAATTC	CAATTCGCGGTGGCCATC	120					
Db	1	ATGAAGATGTGGCATTTCCAAAAGAAATCAAAAACAATG----	AAAAACCGCGTGGCCATC	54					
QY	121	ACCCCGCGCGGGGTGGCGGAATCAACCCGTGTGGCCATGAGTGCT---	CATCCAGGCA	177					
Db	55	ACCTCGCGCAGGGGTGTGATGACGCTCCTCAACAGCGGGGCATGAGTGTGTATGTGGAGACG		114					
QY	178	GGTGCCTGGAGAGGGCTTCGGCTATCAACGACGCGGATTTCAAGCGGCGACGCGGCAACTG		237					
Db	115	GAAAGGCGGGGCTGGGTCGGGGGTTTTCCGATTCGCGATGAAAAAGCCGGGGCAGCTGAT		174					
QY	238	GTTCGGACCGCCGACCAAGTGTGGGCCACGCTGATTTATGTCACAGTCAAGCAAGAACCG		297					
Db	175	CGTTTGGCAACGTGTGAGAAATGCTTTGAGAGCGGGGAGATGTGTTAAAGTGAAGAACCCG		234					
QY	298	ATAGCGCGGAAATACGGCGCGCTGCGACACGCGGACGATCTTGTTACAGTCTTTGCAATTTG		357					
Db	235	CTGGCTTCGAGAGTTCGCGATTTTGCGCCCGGATTTGATTTTGTTAGTATTTGCAATTA		294					
QY	358	GCCGCGTCAAGTGTTCGACCGCATGCGTTTGATTCGGGACCCAGTCAATTCCTAC		417					
Db	295	GCCGCGGCGGAACGCGCTCACGAAAGCCGTCGTCGAGCAAAAAGTGTGTGCGCATGCTTAC		354					
QY	418	GAAACCGTCCAGACCCCGGACGCGGACACTACCCGCGTTCGCGCCGATGAGGAGTGGCC		477					
Db	355	GAAACGCTGAGCTGGCGAAGCGCTGCGCTGCGCATGTTGACGCGGATGAGGAAGTGGCC		414					
QY	478	GATCGACTGCGCCCGCCAGGTTGGCGTTACACACTGATGCGAACCAGGGGGCGCGGT		537					
Db	415	GCGCGCATGTGTGTGAATCGCGCGCCCAATGTTCTTCGAAAGCGGCAAGGGGAAAGGCG		474					
QY	538	GTCGTATGAGGGGGGTGTCGCCGCGCTGCAACCGCCGACGTGTGTGTATGTGGGCGCGG		597					
Db	475	ATTTTGTCTTGGCGCGCTGCCCGGAGTGGCGGCGGCAAGTGAACATATGGGCGGCGGA		534					
QY	598	ACGCGCGGCTCAACAGCAGCCGCGATCGGCAACGCGCATGAGCGCAGCTTACGTTCTA		657					
Db	535	ACGGCGGGACGAAACGCGCGGGAATTCGGGTGCGTCTCGGGTCAAGAGCTGACGATTTTG		594					
QY	658	GACATCAACATCGACAAACTTGGCAACTCGACGCGCCAGTCTGCGCGGATCCACT		717					
Db	595	GACATTTAACCGCCAGCGGCTGCGCGCATCGATGATTTGTGGCGACACGTCAGCGAG		654					
QY	718	CGGTACTCATCGGCTATACGAGCTCGAGGGTGGCGTCAAAAGTGGCAGCTGGTGTAGG		777					
Db	655	CTCATGTCCAACTGTATCCATATGCGCGAGTGGGTGGCGGAATCGGAATTTGGTGTGCTG		714					
QY	778	GCGCTCTGTGTCGAGCGCGCAAGGCAACCAATTAGTCTCAATTACCTTGTGCGCAT		837					
Db	715	GCGCTCTGTATCCCGGGGGGGA---AGGAAAGCTGTGACGGAAGAGATGGTGGCTCG		771					
QY	838	ATGAACCAAGGTGGGTACTGTGTGTGATATAGCATATGACACAGGGGGGCTGTTTGA---		895					
Db	772	ATGAGCGCGGATTCGGTGTGTGTGTCGTAATTCGCATTGACCAAGGCGGCAATTTTCAACG		831					
QY	895	GGCATCAGACCGACCACTACGACACACGAGCTTGCGCGGTGCAGAGACGCTGTATTAC		954					
Db	832	ACCGACCGACCGTACGACGACGACGATCCGACTATCGTCAACAGCGGCTGCTCATTAAC		891					
QY	955	TGGGTGGCGAAATGATGCGCGGTGCGGAAAGAGTGAAGCTACAGAGGCTGACCAAGCGG		1014					
Db	892	GCGCTTGGCAACATTTGCGCGG---GTGCGCGGACGCTGACATTTGCGGTTATCGAAGCTC		948					
QY	1015	ACGATGCCGATATGTCTGAGAGCTTGCCGACCATGCTGCGGGGCGGCTGCCGTGCAAT		1074					

```

RESULT      6
ID           VS2155 standard; DNM: 28171 BP.
AC           VS2155:
DT           23-Oct-1998 (first entry)
DE           Streptococcus pneumoniae genome fragment SEQ ID NO:22.
KW           Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
OS           computer readable medium; vaccine; pharmaceutical composition; ds.
PN           WP0818931.42.
PD           07-MAY-1998.
PF           30-Oct-1997; U19588.
PR           31-Oct-1996; US-029960.
PA           (HUMA) - HUMAN GENOME SCI INC.
PI           Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI           Kunsch CA, Rosen CA.
PI           WPI: 98-272225/24.
PT           Computer-readable medium with recorded Streptococcus pneumoniae
PT           polynucleotide sequences - useful in diagnostic kits and assays, and
PT           pharmaceutical compositions and vaccines for Streptococcus
PT           pneumoniae
PS           Claim 1: Page 273-289; 1409pp; English.
CC           The present invention describes a computer readable medium which has
CC           the nucleotide sequences SEQ ID NO:1 to 391 (VS2134 to VS2524) identical
CC           on it, or a representative fragment of a sequence at least 95% identical
CC           to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC           to 391 (VS2134 to VS2524) are genomic fragments from Streptococcus
CC           pneumoniae. The present invention also describes an isolated nucleic acid
CC           molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC           genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC           by a process comprising: (a) screening a genomic DNA library using as a
CC           probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC           to 391, identifying members of the library which contain sequences
CC           that hybridize to the target sequence and isolating the nucleic acid
CC           molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC           from an organism, amplifying nucleic acid molecules whose nucleotide
CC           sequence is homologous to amplification primers derived from the
CC           fragment of the S.pneumoniae genome to prime the amplification and
CC           isolating the amplified sequences. The computer readable medium can be
CC           used in a computer-based system for identifying fragments of the
CC           S. pneumoniae genome of commercial importance, or expression modulating
CC           fragments of the S. pneumoniae genome. Products from the present
CC           invention can be used in diagnosis kits and assays, and pharmaceutical
CC           compositions and vaccines for S. pneumoniae.
SQ           Sequence 28171 BP; 8507 A; 5080 C; 6440 G; 8144 T;

Query Match      15.9%; Score 198.4; DB 1; Length 28171;
Best Local Similarity 53.4%; Pred. No.3,7e+33;
Matches 516; Conservative 0; Mismatches 431; Indels 19; Gaps 4.

Db      89 AANAACACGAATTCCAATTCGCGTGGCCATACACCCGCGCGCGCGGAGGAACATACCC 148
      21689 AATTTAAATAAAGTAAGAAACCGGTGGCCCTACACCTGAGGTGTTCAATAGCTTAGTTA 21630
      149 GTCTGGCCATGAGCGGTCTATCTCAGACGAGGTGCGGAGAGGGCTCGGATACCGAGC 208
      21629 GTCTGTGTATCTGTCTTATTCGAAACAAATCTGTCTGTCTGTCTGTCTTACGTGATG 21570
      QY 209 CGGATTTCAAGGGGCGGAGCGCCCACTGGTGGGACCGCGGACCGAGGTGGGGCGGAGC 268
      Db 21569 CTGCTGCTATCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 21513
      QY 269 CTGATTTATTTGCTCAAGGTCAAGAAACCGATAGCGGGCGGATTCGCGTCCGACAGC 328
      Db 21512 CAGAGTGTGTTGTGAAGAGTAAAGAAATCTTTAAGTCTGTAATCGGTTACTTCGCGAGC 21453
      QY 329 GGGAGATCTTGTTCAAGTCTTTCATTTTGGCCGCGTCAGAGTCTTACACGATCGCTGT 388
      Db 21452 ATCTTCTTCTTCTTCACTTCTGCAATGAGCGCGCTGCTCCAGATTAAGCAGATGCTATGT 21393

```

QY	389	TGGAATTCGGGCACCCAGCTCAATTGGCTCAGAGACCGTCCAGACGCCGACGGCGACTAC	448
Db	21392	T-----AACAGCAAAAACAACTGAACACTGTTCGTGACAAATACAGACAACACTAC	21345
QY	449	CCCTGCTTGCCCCCATGTAGAGCGAAGTGGCCGGTGCAGCTGCGCCGCCAGGTGGCGCTTACC	508
Db	21344	CGCTCTCGTTCTTCATTAGATGAGTGAGTTGCAGACGCTATGGCTGTTCATAATTCGACGCTCACT	21285
QY	509	ACCTGATGCGAACCAGAAGGGGCGCCGGTGTGCTCATGGGGCGGGGTGGCCGGCGCTCGAAC	568
Db	21284	TCTCTTAACCAAGCTGGTGGCTGTGGTGTCCTACTGGTGGTGTACAGAGTGTTCCAA	21225
QY	569	CGGCCGACGCTGGTGTGATCGGCGCCGCCACCCGCGCTACAAACGACGCGCATCGCCA	628
Db	21224	AAGAAAAAGTAACTATCATACGGTGGTGGTGCTGTGGTACATGCTGCGCCCATCGCCC	21165
QY	629	ACGCGATGGGGCGCCGACCCGTTAGCGTTCTTAGACATCAATCGACAAACTTGGCAACTCG	688
Db	21164	TTGCTCTTGGTGCTCAAGTGACTATTATTAGATATTAGTTTCCAAAGCGTCTCTCAGTTCTAG	21105
QY	689	ACGCGAGTTCGCGCGCCGATCCACACTCGCTACTCATCGGCTACGAGTCTGAGGGTG	748
Db	21104	AAGAGTCTTTGGAAAGTCAAAATTCAAAGCTTATGTCTAATTCATTCAACATTGAAGCAA	21045
QY	749	CCGTCAAACGTGCCGACCTCGTGTGATTTGGGGCCGCTCCTGGTCTCAGGCCCAAGCACCA	78
Db	21044	GGTGAGAGATCTCTATGTGGGATGTGGACCAATTCCTCCCTGGTGGAAAAAGCACCG	20985
QY	809	AATTATGTCGCAATTCACCTGTGCGGCATATCAAAACAGAGTGGGCTGAGTGTGATATAG	868
Db	20984	AATTGTGACATATGATGTGTCAAAACAATGCTGCACAGCTGTCTGA-TGCTTGACGTGG	20926
QY	869	CCATGACACGAGGGCGGCTGTTTGA--AGGCTCACGACCGACACCACTACGACACCGCA	925
Db	20925	CTGTGACCAAGGTGGCGCTTATCGAAACAGCTGACCGCTGTGCACACGACATGAACCCG	20866
QY	926	CCTTGCCCTGACAGCAAGACGCTGTTTACTGGTGGCGCAACATGCCCGCTCGGGTGGCGCA	985
Db	20865	TCTATGAAAAACACGGTGTCTTCCACTATGCGGTTCGCAATATCCTGTGTGGGTGGTCTC	20806
QY	986	AGACGTGACACTACGGCTGCTGACCAACGCGACGATGCGCTATGTGCTCGAGCTTCCGACC	1045
Db	20805	GGACTTCAACCAATCGCCCTTAACCAATGTCATCTTCTTATATCGAAGCTTTGGCTGGCA	20746
QY	1046	ATGGCT 1051	
Db	20745	AAGGAT 20740	
<hr/>			
RESULT 7			
ID	V74442/c		
AC	V74442; standard; DNA; 9280 BP.		
DT	16-MAR-1999 (first entry)		
DE	Staphylococcus aureus contig SEQ ID #131.		
KW	Computer readable medium; vaccine; S.aureus infection; immunodetection;		
KM	cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;		
KV	skin infection; surgical wound infection; scalded skin syndrome;		
KX	toxic shock syndrome; ds		
OS	Staphylococcus aureus.		
FH	Key	Location/Qualifiers	
FT	misc-feature	841..900	
FT	/note-	/tag-a	"these bases represent a line of missing text in the sequence listing to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc-feature	2641..2700	
FT	/tag-b	/note-	"these bases represent a line of missing text in the sequence listing to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT			
FT			
FT			

Query Match	Best Local Similarity	Score	DB 1:	Length	9280;
Matches	484;	Conservative	0;	Mismatches	466;
				Indels	6;
				Gaps	2
157	CATGAGTGCTCATCCAGCAGGTCGGGAGAGGGCTCGGTATACCCAGCCGGATTTC	216			
9280	CATCAGCTTTTACGTGGAACCAATATCGGGTTTCAGATCATCTCTTGAATATGAGATTTC	9221			
217	AAGCGCGCAGCGCGCACTGTCGGACACCCCGACACAGTGTGGCCAGCGTATTTA	276			
9220	AAAGAGCAGGGGCTGAGATTCTTGCTGAACAAGCAAAAGATTGG--GATGTGGATTAG	9164			
277	TTCGCTCAAGGTCAAGAACCGATAGCGGGGAGATACGGCCGCTGGCAGACGGCATC	336			
9163	GTTATTAAAGTTAAACAACCACTTGAATGTGAATATCCATATTTTAAAGAGGCTTGTGA	9104			
337	TGTTTCAGGTTCTTGCAATTTGGCCGGCGTGACGTGCTTGCACCGAGTGGTGTGATTCC	396			
9103	TATATCATCTTACTTATTACCAATGAAGAAAAAATTAACAACACTTGTATAGATAGA	9044			
397	GGCACCACTCATTTGCCATAGAGACCGCTCCAGACCGCCGACGGCGCATATCCCTGCTT	456			

```

Db 9043 AAGTAATTAGTATGATATGAGATGCTGCAATTACCAAGACATTTTACATGTTA 8984
Qy 457 GCCCCGATGACCAAGATGCGCCGTCGATCGCCGCCAGGTTGGCGCTTACCACCTGATG 516
Db 8983 TCACCAATGATGAGTAGACAGAAAGATGTCAGTCAAGTTGGCGAGATTCCTACAA 8924
Qy 517 CGAACCCCAAGGGGCGCGGTGCTGATGGGCGGGGTCGCCGCGTGCAGACCGGCCAGC 576
Db 8923 AACTTATATGTTGATATGAGTAATCTACTAGTGTGTCCAGAGAGTACCTAAGGTTAAA 8864
Qy 577 GTCTGTGATGCGCGCCGCGACCGCGGCTACAAACGACGCGCCGATCGCAAGCGCAT 636
Db 8863 GTACTATATGCTGGTGGTGCAGACAGAAATGCAAGTAAATGCACTAGTACTA 8804
Qy 637 GGGCGACCGCTTACGCTTACGATGATGATGATGATGATGATGATGATGATGATGAT 696
Db 8803 GGTCCAGATGTTACGATTTTATGATGTTATGATGATGATGATGATGATGATGATGAT 8744
Qy 697 TTCTGCGCGCGGATGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 756
Db 8743 TTGCGTGGAGCTGATACATACATATATGCAATGCTGATGATGATGATGATGATGAT 8684
Qy 757 CGTGCGGACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 816
Db 8683 CAAAGTATTTAGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8624
Qy 817 TCGAATTACTGCTGCGCATATGAAACAGGTGGGCTGAGTGGGATGATGATGATGATGAT 876
Db 8623 ACAGACATGATGATTAACAAATGAAATGATGATGATGATGATGATGATGATGATGATG 8564
Qy 877 CAGGCGCGCTGCTGCAAGGCTC--ACGACGACACCTGACGACACCGGCTGCTGCTG 933
Db 8563 CAAAGCGGATTTTGAACAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 8504
Qy 934 GTGACGACGACGCTGTTTACTGCTGCGGCAATGCGCGCTGCTGCTGCTGCTGCTG 993
Db 8503 AAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8444
Qy 994 ACCTACGCGCTGACCAAGCGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1053
Db 8443 ACCTTACCTTAAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8384
Qy 1054 CGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1109
Db 8383 AGAGAGCATTTTAATCAATCAACCATATCATAGTTTAAATCACTTACAAAGG 8328

RESULT 8
ID T67971 standard: DNA: 1074 BP.
AC T67971;
DT 15-JUN-1997 (first entry)
DE H. pylori membrane protein ORF 05cp20518orf61.
KW Vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW membrane; amino acid; metabolism; ds.
OS Helicobacter pylori.
FH key Location/Qualifiers
FT cds 1..1074
FT /*tag- a
FT /note- "no stop codon given"
PN MO9640893-A1.
PD 19-DEC-1996.
PE 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR ) ASTRA AB.
PI Berglundh O, Smith D, Mellgaerd BL;
DR WPI: 97-052306/05.
DR P-PSDB: W20718.
PT Helicobacter pylori nucleic acid sequences and related
polyptide(s) - useful for vaccines to treat or prevent H. pylori

```

```

PT infection, and to detect Helicobacter
PS Claim 27: Page 819; 1481p; English.
CC The present sequence encodes a Helicobacter pylori membrane
CC protein likely to contain four membrane spanning regions.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 53679) was determined from
CC overlapping contigs generated by mechanically shearing the
CC bacterial DNA. The sequences were analysed for ORF of at least 180
CC nucleotides, and the predicted coding regions defined by computer
CC evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences
CC of interest, particular regions can be isolated from H. pylori by
CC PCR amplification for recombinant polypeptide production, e.g. in
CC E. coli hosts.
SQ Sequence 1074 BP: 302 A; 189 C; 297 G; 286 T;

```

```

Query Match 7.68; Score 94.8; DB 1; Length 1074;
Best Local Similarity 46.18; Pred. No. 1.1e-11;
Matches 370; Conservative 0; Mismatches 417; Indels 15; Gaps 1;

```

```

Qy 271 GATTATTTGCTCAAGCTCAAGAACCGATAGCGCGGAAATCGCGCCCTGCGACAGCG 330
Db 133 GATTGTTGGTCAATGCAAGAGCTTTAGACGATGATCCCTTTGCTCAAGAAAA 192
Qy 331 CAGATCTTGTCAAGTCTTCTGATTTGGCGCGTACGCTGCTGACGATGCTGCTG 390
Db 193 GCGACTGCTGTTATTTATTTGATTTAGCGTATCAAAAAACCTTGTTGAATTTTATT 252
Qy 391 GATTCGGGACACGCTAATTCCTACAGACCGTCCAGACCGCGGACGCTACACC 450
Db 253 AATAAAAAATCACTTCTTATTTGCACTGAACCATTTGCGGCGCTTAAACGATACCT 312
Qy 451 CTGCTGCGCGGATGAGGAGTGGCGGTGATGCTGCGCGCGCGCGCGCGCGCGCGCG 502
Db 313 ATTTAGCGCTTATGAGGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372
Qy 502 -----GCTTACCACTGATGCGAACCCAAAGGGCGCGGTGCTGATGAGGGGGTG 555
Db 373 TTACTGGTTTAAAGCATGTTAAAGGTTTATGGGTAAGGGGTCATGCTAGGGGGTTTG 432
Qy 556 CCGGCGTGAACCGCGGAGCTGCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 615
Db 433 TCGGTTGGCAAGGCGTAAATCGCTGATTTGAGCGCGGTGTTGGCATGAGAGC 492
Qy 616 GCGCGCATCGGCAAGCATGAGGCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 675
Db 493 GCGAAGTCTTAAAGCCAAATGCGGCTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 552
Qy 676 CTGCGCACTGCAAGCGGAGTCTTGGCGCGGATCCACACTGCTACTGAGGCTAC 735
Db 553 TTACAAAAACACCTTATATCATTTTATGATTTAGAAAGCTTAAAGCGTAAAGAAC 612
Qy 736 GAGCTCGAGGCTGCGTCAAAAGTGGACCTGATTTGGGCGCTGCTGCTGCTGCTG 795
Db 613 AATATCATTAAGCCTTAAAGCGGCGGTAGTGGGCGGTGCTGCTGCTGCTGCTGCTG 672
Qy 796 GCCAAGGACCCCAATTTGCTGCAATTCATTTGCGGCGATATGAAACGAGTGGCGTA 855
Db 673 AGCCAAACCCCTTAAAGTATCTTAAAGAGCATTTAAATCATGACGACCAAGGGGTA 732
Qy 856 CTGTTGATATAGCATGACAGAGGGCGCTGTTGCAAGGCTCAGCAGCAGCAGCATAC 915
Db 733 GTCATTTGATGCTGCTGAGTTTAGGGGGGTCATTAAGCATACGCAAGCAAGCAT 792
Qy 916 GACCAACCGAGTTCGCGCTGACGACAGCTGTTTACTGCTGCTGCTGCTGCTGCTG 975
Db 793 TCTAACCGGCTGATGTTGAGAAAGTGTGTTGATTTATGCTGCTGCTGCTGCTGCTG 852

```

QY 976 TCGGTGCGGAGACGTGACCTACGCGCTGACCAAGCGACGATGCTATGCTGAG 1035  
 DB 853 ATTGTCCCTAAAGAGACTTACGCGCTTATAGCATGAGTGTGCTATTTGTGTAT 912  
 QY 1036 CTTCGCGACCATGCTGCGGG 1057  
 DB 913 TATTTAGGATGCTTGAAGG 934

RESULT 9  
 V43039/c  
 ID V43039 standard; DNA: 544 BP.  
 AC V43039;  
 DT 09-NOV-1998 (first entry)  
 DE Streptococcus pneumoniae polypeptide coding region.  
 KW Polypeptide; ORF: open reading frame; infection; bacterial;  
 OS Streptococcus pneumoniae; diagnosis; prophylaxis; ds.  
 FH Key Location/Qualifiers  
 FT CDS complement (59..334)  
 FT /\*tag- a  
 FT /note- "polypeptide"  
 PN MO9823631-A1.  
 PD 04-JUN-1998.  
 PE 24-NOV-1997; U21976.  
 PF 27-NOV-1996; U5-031879.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PI Black MT, Hodgson JE, Knowles DDC, Lonetto MA, Nicholas RO,  
 PI Reid RH, Zarfos PN;  
 PI WPI; 98-322654/28.  
 DR P-PSDB; W62769.  
 DT Streptococcus pneumoniae polynucleotides - useful for developing  
 PT products for diagnosis, prevention and treatment of infections e.g.  
 PT pneumonia, bacteremia, meningitis or endocarditis  
 PS Claim 1, Page 165-166; 181pp; English.  
 CC The sequence is that of a Streptococcus polypeptide coding region.  
 CC The polypeptide can potentially be used for the diagnosis and  
 CC prevention of bacterial infections, especially sp infection.  
 CC It may be used for the treatment of diseases such as otitis media,  
 CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural  
 CC empyema, endocarditis or infection of the cerebrospinal fluid.  
 SQ Sequence 544 BP; 149 A; 121 C; 124 G; 150 T;

Query Match 5.6%; Score 69.2; DB 1; Length 544;  
 Best Local Similarity 53.4%; Pred. No. 2.3e-06;  
 Matches 191; Conservative 0; Mismatches 163; Indels 4; Gaps 2;

QY 697 TTCGCGCGCGATCCACACATCGCTACCTACGCGCTACGAGCTCGAGGTCGCGTCAA 756  
 DB 543 TTGGAATCAATCAATCAATTTATATATATCAATCAATGAGCAAGATATGGA 484  
 QY 757 CGTGCAGCAGCTGATGATGGGCGCTCGTGCAGGCGCAAGCAACCAATTAATGTC 816  
 DB 483 GATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 424  
 QY 817 TCGAATTCATTTGCGCGCATATGAAACAGTGGCGTATGATGATGATGATGATG 876  
 DB 423 ACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 365  
 QY 877 CAGGCGCGCTTTTGA---AGGCTCAGCAGCGACACCTTCACACCGAGTGTGCC 933  
 DB 364 CAGGCGCGCTTTTGA---AGGCTCAGCAGCGACACCTTCACACCGAGTGTGCC 305  
 QY 934 GTGACGACAGCAGCTGTTTACTGCTGCGGCAACATGCGCGCTGCGGCGCAAGCTCG 993  
 DB 304 AAAGCAGGCTGTTTCTGCACTATGCGCGTGCATATGCTGCGCGCTGCGCGCTCA 245  
 QY 994 ACCTACGCGCTGACCAAGCAAGCAGTGCCTATGCTGCGAGCTTGCAGCAGTGCCT 1051  
 DB 244 ACCATCGCGCTTAAACCAAGTCACTCTCTTATATGAGAGCTTTGGCTGGCAAGAT 187

RESULT 10  
 ID V74370 standard; DNA: 31096 BP.  
 AC V74370;  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SEQ ID #59.  
 KW Computer readable medium; vaccine; S. aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scaled skin syndrome;  
 KW toxic shock syndrome; ds.  
 OS Staphylococcus aureus.  
 FH Key Location/Qualifiers  
 FT misc\_feature 1201..1260  
 FT /\*tag- a  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc\_feature 3001..3060  
 FT /\*tag- b  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc\_feature 4801..4860  
 FT /\*tag- c  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc\_feature 6601..6660  
 FT /\*tag- d  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc\_feature 8401..8460  
 FT /\*tag- e  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc\_feature 10201..10260  
 FT /\*tag- f  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc\_feature 12001..12060  
 FT /\*tag- g  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc\_feature 13801..13860  
 FT /\*tag- h  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc\_feature 15601..15660  
 FT /\*tag- i  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc\_feature 17401..17460  
 FT /\*tag- j  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"













GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:58:42 ; Search time 341.15 Seconds  
(without alignments)  
474.370 Million cell updates/sec

Title: US-09-362-485-2  
Perfect score: 1245  
Sequence: 1 ATCTGCAGATATATCGAAC.....AAGCGATGATGCGCCGCG 1245

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/PTCUTS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/Dockfiles1.seq:\*  
  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47.6	3.8	1620	3	US-08-461-775-10 Sequence 10, Appl
2	47.6	3.8	28958	1	US-08-258-261B-6 Sequence 6, Appl1
3	47.6	3.8	28958	1	US-08-456-837-6 Sequence 6, Appl1
4	47.6	3.8	28958	1	US-08-457-342-6 Sequence 6, Appl1
5	47.6	3.8	28958	1	US-08-457-646A-6 Sequence 6, Appl1
6	47.6	3.8	28958	2	US-08-458-076A-6 Sequence 6, Appl1
7	47.6	3.8	28958	2	US-08-764-233A-4 Sequence 6, Appl1
8	47.6	3.8	28958	2	US-08-457-335A-6 Sequence 6, Appl1
9	47.6	3.8	28958	2	US-08-729-214-6 Sequence 6, Appl1
10	47.6	3.8	49377	2	US-08-816-105A-2 Sequence 1, Appl1
11	46	3.7	2582	4	US-08-816-105A-2 Sequence 2, Appl1
12	45.4	3.6	833	3	US-08-403-852D-3 Sequence 3, Appl1
13	45.4	3.6	5392	3	US-08-403-852D-1 Sequence 1, Appl1
14	42.4	3.4	1833	3	US-08-403-852D-6 Sequence 6, Appl1
15	42.4	3.4	2185	1	US-08-173-508-3 Sequence 3, Appl1
16	42.4	3.4	2185	3	US-08-265-310-3 Sequence 3, Appl1
17	41.2	3.3	2414	7	US-08-265-310-3 Patent No. 5248599
18	41.2	3.3	2668	3	US-08-461-775-11 Sequence 11, Appl1
19	41	3.3	20235	2	US-07-642-734C-3 Sequence 3, Appl1
20	40.8	3.3	44377	3	US-08-804-227C-7 Sequence 7, Appl1
21	40.8	3.3	44377	4	US-08-804-198-1 Sequence 1, Appl1
22	40.6	3.3	4267	3	US-08-690-473-1 Sequence 1, Appl1
23	40.6	3.3	13001	2	US-08-458-568A-11 Sequence 11, Appl1
24	40.2	3.2	2064	1	US-08-343-428-1 Sequence 1, Appl1
25	40.2	3.2	11219	2	US-07-642-734C-1 Sequence 1, Appl1
26	39.6	3.2	474	3	US-08-403-852D-14 Sequence 14, Appl1
27	39.4	3.2	459	4	US-08-387-942C-35 Sequence 35, Appl1

C	28	39.4	3.2	1998	7	5212296-8	Patent No. 5212296
C	29	39.4	3.2	12588	4	US-08-387-942C-1	Sequence 1, Appl1
C	30	39.4	3.1	420	1	US-08-470-179-148	Sequence 148, App
C	31	39.2	3.1	2353	6	PCT-US92-06840-1	Sequence 1, Appl1
C	32	39	3.1	43280	3	US-08-804-227C-1	Sequence 1, Appl1
C	33	38.8	3.1	1524	7	5229279-1	Patent No. 5229279
C	34	38.8	3.1	1525	7	5229279-1	Patent No. 5229279
C	35	38.8	3.1	8051	4	US-08-576-526A-2	Sequence 2, Appl1
C	36	38.8	3.1	11219	2	US-07-642-734C-1	Sequence 2, Appl1
C	37	38.6	3.1	1187	2	US-08-440-856A-2	Sequence 2, Appl1
C	38	38.6	3.1	3231	2	US-08-074-121-4	Sequence 4, Appl1
C	39	38.6	3.1	3231	6	PCT-US94-06447-4	Sequence 4, Appl1
C	40	38.4	3.1	1215	4	US-08-947-726A-1	Sequence 20, Appl1
C	41	38.2	3.1	2109	4	US-08-555-568B-20	Sequence 1, Appl1
C	42	38	3.1	30001	1	US-08-125-468-1	Sequence 1, Appl1
C	43	38	3.1	30001	3	US-08-474-933-1	Sequence 1, Appl1
C	44	37.8	3.0	2588	3	US-08-796-414B-6	Sequence 6, Appl1
C	45	37.8	3.0	15664	1	US-08-402-282-3	Sequence 3, Appl1

## ALIGNMENTS

RESULT 1  
US-08-461-775-10  
; Sequence 10, Application US/08461775  
; Patent No. 5858773  
GENERAL INFORMATION:  
APPLICANT: MAZODIER, Philippe  
APPLICANT: GUGLIEMI, Gerard  
TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE  
TITLE OF INVENTION: INITIATION OF TRANSCRIPTION  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,775  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/050,313  
FILING DATE: 10-MAY-1993  
APPLICATION NUMBER: FR 9011186  
FILING DATE: 10-SEP-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 010830-035  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1620 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1620  
US-08-461-775-10

Query Match 3.8%; Score 47.6; DB 3; Length 1620;  
 Best Local Similarity 45.7%; Pred. No. 0.0031;  
 Matches 321; Conservative 0; Mismatches 369; Indels 12; Gaps 4;

120 CACCCGCGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 179  
 879 CACCTTCACCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 938  
 180 TCGCGAGAGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 239  
 939 TCGGAGAGTGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 988  
 240 CCGGAGAGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 299  
 999 GAGGAGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 1058  
 300 AGCGGAGAGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 359  
 1059 CAGTGCAGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 1118  
 360 C-GCGTACGCTGTTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 418  
 1119 CCGGAGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 1178  
 419 AGACGCTGCGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 476  
 1179 GCACGCTGCGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 1238  
 477 CCGTGCAGTGCCTGCGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 533  
 1239 CCGTGCAGTGCCTGCGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 1298  
 534 CCGTGCAGTGCCTGCGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 591  
 1299 CCGGAGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 1358  
 591 ---GCGGAGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 647  
 1359 GATGCGGAGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 1418  
 648 TACGCTTACGATCAACATGAGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 707  
 1419 CAAGGCGGAGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 1478  
 708 GATGCGGAGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 767  
 1479 CATGCGGAGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 1538  
 768 GGTGATGCGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 809  
 1539 GGTGATGCGGCGGTGCGGACTACCCCTGTCGATGAGTGTCTATCCAGGCGAG 1580

RESULT 2  
 US-08-258-261B-6  
 Sequence 6, Application US/08258261B  
 Patent No. 563949

GENERAL INFORMATION:  
 APPLICANT: Schupp, Thomas  
 APPLICANT: Ligon, James M.  
 APPLICANT: Beck, James Joseph  
 APPLICANT: Hill, Dwight Steven  
 APPLICANT: Ryals, John Andrew  
 APPLICANT: Gaffney, Thomas Deane  
 APPLICANT: Lam, Stephen Ting  
 APPLICANT: Hammer, Phillip E.  
 APPLICANT: Urnes, Scott Joseph  
 TITLE OF INVENTION: Genes for the syntheses of  
 TITLE OF INVENTION: antipathogenic substances of  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Ciba-Geigy Corporation

STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10532  
 COMPUTER READABLE FORM:  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION NUMBER: US/08/258, 261B  
 FILING DATE: 08-JUN-1994  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/457, 205  
 FILING DATE: 01-JUN-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Elmer, James Scott  
 REGISTRATION NUMBER: 36,129  
 REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-541-8614  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 28958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-258-261B-6

Query Match 3.8%; Score 47.6; DB 1; Length 28958;  
 Best Local Similarity 44.1%; Pred. No. 0.0076;  
 Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

154 GCGCATGAGG1JCTCATCCAGGAGTGGCGGAGGCGGTGCTATCCAGGCGAG 213  
 25306 GCGCATGAGGAGTGGCGGAGGCGGTGCTATCCAGGCGAG 25365  
 214 TTCAAGGCGGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAG 273  
 25366 TTGAGAGGCGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAG 25425  
 274 TTATGCTCAGGTCAGGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAG 333  
 25426 TTGCTGCGGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAG 25485  
 334 ATCTGCTCAGGTCGATGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAG 393  
 25486 AAGCTGCTTCTGCTATCCAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAG 25445  
 394 TCGGCGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAG 453  
 25546 CTGAGCTCTGCGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAG 25605  
 454 CTGAGCTCTGCGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAG 511  
 25606 CAGCTCAGGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAG 25665  
 512 TGATGCGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAG 571  
 25666 CCGCTGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAG 25725  
 572 CCGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAG 630  
 25726 CGCTGCTGCGGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAG 25785  
 631 GCGATGCGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAGGCGGTGCTATCCAGGCGAG 690

DB 25786 GCGTGTGCGAGGCGCTCTCTCCCTCGAGAGAGCGGCGCCCGCATCGCCGCTCGGCGAGC 25845  
 QY 691 GCCGAGTTCTGGGCGCGGATCCACACTGCTACTCATGCGCCCTAGAGCTCGAGGTGCG 750  
 DB 25846 AAGGCCCTACACCGCTCGCGGCAAGGCGCCATGCGCCGCTCGAGCTCGGCGCTCC 25905  
 QY 751 GTCGAACGTGCGCCTGTGATTGGGCGCCCTGCTGTCAGGCGCC 798  
 DB 25906 GACCTCAGACCTACTCCTGCTCCCTGGGGGAGAGGCTCTCCATCGCC 25953

## RESULT 3

US-08-456-837-6  
 ; Sequence 6, Application US/08456837  
 ; Patent No. 5643774

## GENERAL INFORMATION:

APPLICANT: Schupp, Thomas  
 APPLICANT: Ligon, James M.  
 APPLICANT: Beck, James Joseph  
 APPLICANT: Hill, Dwight Steven  
 APPLICANT: Ryals, John Andrew  
 APPLICANT: Gaffney, Thomas Deane  
 APPLICANT: Lam, Stephen Ting  
 APPLICANT: Hammer, Phillip E.  
 APPLICANT: Uknes, Scott Joseph  
 TITLE OF INVENTION: Genes for the synthesis o  
 TITLE OF INVENTION: antipathogenic substances  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ciba-Geigy Corporation  
 STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10532

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/456,837  
 FILING DATE: 01-JUN-1995  
 CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/457,205  
 FILING DATE: 01-JUN-1995  
 APPLICATION NUMBER: 08/258,261  
 FILING DATE: 08-JUN-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Elmer, James Scott  
 REGISTRATION NUMBER: 36,129  
 REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-541-8614  
 TELEFAX: 919-541-8689

## INFORMATION FOR SEQ ID NO:

6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 28958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-456-837-6

Query Match 3.8%; Score 47.6; DB 1; Length 28958;  
 Best Local Similarity 44.1%; Pred. No. 0.0076;  
 Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

QY 154 GCGCATAGGTGCTCTCATCCAGGAGTGCOCGAGAGGGCTCGGCTATCAACGAGCGGAT 213  
 DB 25306 GCCCAGCAGACCTCACCTCTGCGGATGTGGCTATTGCTGGCCACACCGCGCCAC 25365  
 QY 214 TTCAGGCGGAGGCGGCGCAACTGTGTGCGGCGCGGAGGCTGTGGCCGAGCTGAT 273  
 DB 25366 TTCAGGAGCGGCGGCGCTGTGTGAGGCCACACCGGAGAGGCTCTCTCGGCGTGCAC 25425  
 QY 274 TTATGTCTCAAGTCAAGAAGACCGATAGCGCGGATATAGCGCGCTGTGACAGCGGACG 333  
 DB 25426 TCGCTCCCGCAGACAGCCCGCCGAGCAGCGTCTCTGAGAGGAGGAGAACCGAGCG 25485  
 QY 334 ATCTGTTCAGCTTCTGCAATTGCGCGGCTGACGTCTGACCGATGCTGTGAT 393  
 DB 25486 AAGCTGCTTGTCTTCTTCTGCGCAAGCTTCCAGTGGAGAGATGGCTCTGTG 25545  
 QY 394 TCCGCAACACGCTCAATTGCACGAGACCGTCCAGACCGGCGACCTACCCCTG 453  
 DB 25546 CTCGACTCTCTGCGCGCTTCCGCGCTCAGTCTGACGATGACGAGCGCGCTCGCTCT 25605  
 QY 454 CTTGCCCCGATGAGCGAAGTGGCGGCTGAGTGGCGCCCGCAGCTTGGCGCTTACGAC--C 511  
 DB 25606 CACGTGAGTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25665  
 QY 512 TGATCGAACCAGAGGGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 571  
 DB 25666 CGCTGCGAGCTGTACAGCCCGCCTCTTCCGCTGATGCTCTCTGCGGCGCTCTG 25725  
 QY 572 CCGAGC-TGCTGTGATGAGGCGCGGCGCGGCGCTTCAACAGCAGCCCGCATGCCAAC 630  
 DB 25726 CGCTGCTGCGGCTGAGAGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 25785  
 QY 631 GGCATGGCGCGGCGGCTTACGCTTCTACATCAACATCGAACAACCTTGGCAACTGCGAC 690  
 DB 25786 GCGTGTGCGAGGCGCTCTCTCTCGAGAGCGGCGCGCGCTGCTGCTGCTGCTGCTGCTG 25845  
 QY 691 GCCGAGTTCTGGGCGCGGATCCACACTGCTACTGCTTACAGGCTTACAGGCTGAGGATGCG 750  
 DB 25846 AAGGCCCTACACCGCTCGCGGCAAGGCGCCATGCGCGCGCTGAGACTCGGCGCTCC 25905  
 QY 751 GTCGAACGTGCGCCTGTGATTGGGCGCCCTGCTGTCAGGCGCC 798  
 DB 25906 GACCTCAGACCTACTCCTGCTCCCTGGGGGAGAGGCTCTCCATCGCC 25953

## RESULT 4

US-08-457-342-6  
 ; Sequence 6, Application US/08.57342  
 ; Patent No. 5662898

## GENERAL INFORMATION:

APPLICANT: Schupp, Thomas  
 APPLICANT: Ligon, James M.  
 APPLICANT: Beck, James Joseph  
 APPLICANT: Hill, Dwight Steven  
 APPLICANT: Ryals, John Andrew  
 APPLICANT: Gaffney, Thomas Deane  
 APPLICANT: Lam, Stephen Ting  
 APPLICANT: Hammer, Phillip E.  
 APPLICANT: Uknes, Scott Joseph  
 TITLE OF INVENTION: Genes for the synthesis of  
 TITLE OF INVENTION: antipathogenic substances  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ciba-Geigy Corporation  
 STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10532

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS



QY 274 TTATGCTCAAGTCAAGAACGATAGCGCGGATATACGGCCGCTTCGCACAGGGCAG 333  
DB 25426 TCGCTCCGACGAGCAGCCCGCCGAGCAGCCTCTCCGAGCAGGAGGACGACGGC 25485  
QY 334 ATCTGTTACAGTCTTGCATATTTGGCCGCGACGCTGCTTGCACGATGTTGGAT 393  
DB 25486 AAGTCTCTCTTCTTCTTCCGCGCAAGGCTCGAGAGGAGGAGGAGGAGGAGGAGG 25545  
QY 394 TCCGACACACGATGATGCTTACGATGAGACCGTCCAGACCGCCGACGAGCTACCCCTG 453  
DB 25546 CTCGACTCTCTCCGCTTCTCCGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTC 25605  
QY 454 CTTCGCGGATGAGCAGGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 511  
DB 25606 CAGCTCGAGTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25665  
QY 512 TGATGCGAACCCAGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 571  
DB 25666 CGCTCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25725  
QY 572 CCGAGC-TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 630  
DB 25726 CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25785  
QY 631 GCGATGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 690  
DB 25786 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25845  
QY 691 GCGAGTCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 750  
DB 25846 AAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25905  
QY 751 GTCAAGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 798  
DB 25906 GACCTCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25953

RESULT 6  
US-08-458-076A-6  
Sequence 6, Application US/08458076A  
Patent No. 5698425  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James M.  
APPLICANT: Beck, James Joseph  
APPLICANT: Hill, Dwight Steven  
APPLICANT: Ryals, John Andrew  
APPLICANT: Gaffney, Thomas Deane  
APPLICANT: Lam, Stephen Ting  
APPLICANT: Hammer, Phillip E.  
APPLICANT: Ukner, Scott Joseph  
TITLE OF INVENTION: Genes for the synthesis of  
antipathogenic substances  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
GEMFIRM REABABLE PGM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,076A  
FILING DATE: 01-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/457,205

FILING DATE: 01-JUN-1995  
APPLICATION NUMBER: 08/258,261  
FILING DATE: 08-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28958 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-458-076A-6

Query Match 3.8%; Score 47.6; DB 2; Length 28958;  
Best Local Similarity 44.1%; Pred. No. 0.0076;  
Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

QY 154 GCGATGAGTGTCTCATCCAGCGAGTGGCGGAGAGGCTGCGCTATCAACGAGCGGAT 213  
DB 25306 GCGCAGCAGCAGCTCACCTCGCGGATGCGCTATTCGCTGCGCAGCAGCGCGCCAG 25365  
QY 214 TTCAGCGCGCAGCGCGCAGCTGCGCGCAGCGCGCAGCGCGCAGCGCGCAGCGCGAT 273  
DB 25366 TTCGAGC AGCGCGCGCTCTGTA3CCCAACCGGAGAGAGCTCTCTCGCGGCTCGAC 25425  
QY 274 TTATGCT CAAGTCAAGAACGATAGCGCGGATAGCGCGCTGCGACAGGGCAG 333  
DB 25426 TCGCTCGACGAGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25485  
QY 334 ATCTGTTACAGTCTTGCAT TTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 393  
DB 25486 AAGTCTCTTCTGCTTCTTCTGCGCAAGGCTGCGAGTGAAGATGAGGAGGAGGAGGAGG 25545  
QY 394 TCCGACACACGATGATGCTTACGATGAGACCGTCCAGACCGCCGACGAGCTACCCCTG 453  
DB 25546 CTCGACTCTCTCCGCTTCTCCGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 25605  
QY 454 CTTCGCGGATGAGCAGGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 511  
DB 25606 CAGCTCGAGTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25665  
QY 512 TGATGCGAACCCAGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 571  
DB 25666 CGGCTCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25725  
QY 572 CCGAGC-TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 630  
DB 25726 CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25785  
QY 631 GCGATGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 690  
DB 25786 GCGTTCGCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25845  
QY 691 GCGAGTCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 750  
DB 25846 AAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25905  
QY 751 GTCAAGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 798  
DB 25906 GACCTCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25953

RESULT 7  
US-08-764-233A-4



Sequence 4, Application us/08764233A  
Patent No. 5716849

## GENERAL INFORMATION:

APPLICANT: Ligon, James M.  
APPLICANT: Schupp, Thomas  
APPLICANT: Beck, James J.  
APPLICANT: Hill, Dwight S.  
APPLICANT: Neff, Snezana  
APPLICANT: Ryals, John A.  
TITLE OF INVENTION: Genes for The Biosynthesis of Sorphen  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 520 White Plains Road, P.O. Box 2005  
CITY: Tarrytown  
STATE: NY  
COUNTRY: USA  
ZIP: 10591

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764,233A  
FILING DATE:  
CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/729,214  
FILING DATE: 09-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/258,261  
FILING DATE: 08-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: 1506/CIP6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 541-8587  
TELEFAX: (919) 541-8689

## INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
LENGTH: 28958 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Sorangium cellulosum  
IMMEDIATE SOURCE:  
CLONE: p98/1  
US-08-764-233A-4

## Query Match

Best Local Similarity 3.8%; Score 47.6; DB 2; Length 28958;  
Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

QY 154 GCGCATGAGGTGCTCATCCAGAGAGTGGCGGAGAGGGCTTCATCCAGCGAGGAT 213  
DB 25306 GCCCAGACGACCTCCCTCGGATGAGGCTTTCCTGCGCCACACCGCGCCAC 25365  
QY 214 TTCAGAGCGGACGCGCAACTGTGCGACCGCCGACAGGATGAGGCGCGAGT 273  
DB 25366 TTGAGACACCGCGCGCTCTCGTAGCCACAAACCGGAGAGGTCTCTTCGCGGTGAC 25425  
QY 274 TTATGCTCAAGTCAAGAACGATAGCGCGGATAGCGCGCTGGGACACGGGCGAC 333  
DB 25426 TCGCTGCCCAAGCAAGCCCGCCGAGCAGCGTCTCGAGGAGGAGCGAACCGAC 25485  
QY 334 ATCTGTTCACGTTCTGATTTGGCGCGGTGACGTCCTTGACACGATGCGTTGTGAT 393  
DB 25486 AAGCTGCTTCGCTTCCTGCGCAAGGCTCGCAGTGGAGAGGATGGCCCTCTGCTG 25545

QY 394 TCCGGACACGCTCAATTCCTACGAGACCGTCCAGACCGCCGAGCGGACACTACCTG 453  
DB 25546 CTCGACTCCTCGCCCGCTTCCCGCTCAGCTCGAAGATGAGACGCGCCCTGCTCT 25605  
QY 454 CTTCGCCGATAGGAGAGTGGCGGATGAGTGGCGGACGATGGGCTTACAC--C 511  
DB 25606 CAGTGCAGTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25665  
QY 512 TGTGCGAACCCAGGGGCGCGGCTGCTGATGAGGCGGGGTGCGCGGCTGCAACGG 571  
DB 25666 CGGTGAGAGGTGTACAGCCCGCCCTTTCGCTGATGCTGCTGCTGCTGCTG 25725  
QY 572 CCGACG-TGTGCTATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 630  
DB 25726 CGCTGCTGCGCTAGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25785  
QY 631 GCGATGGGCGGACCGCTTACGATTCAGATCAACATCGACAACTTGGCACTGAC 690  
DB 25786 GCCTGTGCGCAGGCGCTCTCTCGAGAGCGCGCGCGCGCGCGCGCGCGCG 25845  
QY 691 GCGGATTCGGCGCGCGGATCCACTGCTCTATCGGCTTACGAGGCTGAGGGTGC 750  
DB 25846 AAGCGCTACACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25905  
QY 751 GTCAACGTGCGGACCTGATGATGGGCGCGCTGCTGCGCGCGCGCGCGCG 798  
DB 25906 GACCTCGACACTACCTCGCTCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 25953

## RESULT 8

US-08-457-335A-6  
Sequence 6, Application us/08457335A  
Patent No. 5723759

## GENERAL INFORMATION:

APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James M.  
APPLICANT: Beck, James Joseph  
APPLICANT: Hill, Dwight Steven  
APPLICANT: Ryals, John Andrew  
APPLICANT: Gaffney, Thomas Deane  
APPLICANT: Lam, Stephen Tang  
APPLICANT: Hammer, Phillip E.  
TITLE OF INVENTION: Genes for the synthesis of  
TITLE OF INVENTION: antipathogenic substances  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,335A  
FILING DATE: 01-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/457,205  
FILING DATE: 01-JUN-1995  
APPLICATION NUMBER: 08/258,261  
FILING DATE: 08-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
TELECOMMUNICATION INFORMATION:

```

; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Hammer, Phillip E.
; APPLICANT: van Pee, Karl-Heinz

```

QY	154	GGCCATGAGAGTCTCATCCAGGCAGAGTGGCCGAGAGGGGTGGGCTATACCGACCGGAT	213
Db	25306	GGCCACGAGACCTTACCTCGGGATGAGCTATTGGCTGGCCACACCCGGCCAC	25365
QY	214	TTCAAGGGCGGACGCCUCCAAATGGTGTCTTCTTCCGACACAGAGTGTGGCCAGCCTGAT	273
Db	25366	TTTCGACACACCGGCGCCCTCTCTGTATCCACACCGGAGAGAGCTCTCTCGCGCTCGAC	25425
QY	274	TTATTGCTCAAGGTCAAAAGAACCGATTAGCGGGGAATACGGCCGCTCGCACAGGGGAG	333
Db	25426	TGCGTGGCCAGACAGACGGCCGCCGAGACCGTCTCTGGACGGAGGGAGGCCACGGC	25485
QY	334	AACCTGTTACGTTCTTTCGATTGTGGCCGCGTACGTGCTTGACACCGATGCGTTGGAT	393
Db	25486	AAGCTCGTCTTGTGTTCTTCTGGGCAAGCGCTCGCAGTGGAAAGGATGGCCCTCTCGCG	25545
QY	394	TCCGGACACAGCTCAATTTCCTTACGAGACCGCTCCAGACCGCGGAGGCGGCACTACCCCTG	453
Db	25546	CTCGATCTCTCGCCGCTCTTCGGCGCTCAGCTCGAAGCATGCGAGCGCGCTCGCTCT	25605
QY	454	CTTGCCCCCATGAGCGAAGTTCGCGGTCCGCTACCTCGCGCCCAAGATTGGCGCTTACCAC--C	511
Db	25606	CACGTGAGATGAGAGCGCTGCTCGCCGCTCCGCGCGAGAGAGAGGGGGCCCCCTCCCTCGAC	25665
QY	512	TGATCGGAACCAAGGGGCGCGGCTGTGCTATGAGGCGGGGTGCCCGCGCTCGAACGGG	571
Db	25666	CGCGTGCAGCTGTACAGCGCCCTCTTCTTGGCGTATAGTCTCCCTGCGGCGCCCTTGG	25725
QY	572	CCGAGC-TGCTGTATCGGCGCCGCGACCGCGGCTACAGCAGCACGCCGATTCGCCAAC	630

```

Db 25726 GCCTCGCTCGGAGTGTAGAGCCCGCCGCGTCGCGGCAACGTGACGAGGGAGATGCGCGCC 25785
Oy 631 GGCATGGGCGCGACCGCTTACGGTTCTAGACATCAACATGCAAACTTGGCAACTCGAC 690
Db 25786 GCGTTCGTCGCGAGGGGCTCTCTCCCTCGAGAGCGGGCCCGCATGCGCCCTCGCGACG 25845
Oy 691 GCCGAGTTCGCGGCGCGATCCACATCGCTACTCATCGGCGCTACGAGCTCGAGGGGTGC 750
Db 25846 AAAGCGCTTACCACCGCTGCGCCGCGCAACGAGGCGCATGCGCGCGCTGAGCTCGCGCCCTCC 25905
Oy 751 GTCAACAGTGGCGGACCTGTGATTGGGGCGGCTGCTGTCGAGGCGCC 798
Db 25906 GACGTCGACAGCTTACCTCGCTCCCTGCGGCGAGACAGGCTCTCCATCGCC 25953

RESULT 10
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
GENERAL INFORMATION:
APPLICANT: Ligon, James M.
APPLICANT: Schupp, Thomas
APPLICANT: Beck, James J.
APPLICANT: Hill, Dwight S.
APPLICANT: Neff, Snezana
APPLICANT: Ryals, John A.
TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,233A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/729,214
FILING DATE: 09-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Melgs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: 1506/CIP6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4937 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Sorangium cellulosum
IMMEDIATE SOURCE:
CLONE: p98/1, pJ13, and pVKM15
FEATURE:
NAME/KEY: misc_feature
LOCATION: 383..760
OTHER INFORMATION: /product= "SorP"
OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous

```

```

OTHER INFORMATION: the reductase domains of type I PRSs such as eryA from
OTHER INFORMATION: Saccharopolyspora erythraea."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 927..19874
OTHER INFORMATION: /product= "SorA"
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PRS t
OTHER INFORMATION: are known to be involved in the synthesis of polyketide
OTHER INFORMATION: compounds."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 942..7115
OTHER INFORMATION: /product= "Module 1 of SorA"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7203..12884
OTHER INFORMATION: /product= "Module 2 of SorA"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 13455..19616
OTHER INFORMATION: /product= "Module 3 of SorA"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19871..46318
OTHER INFORMATION: /product= "SorB"
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PRS, ge
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19870..24556
OTHER INFORMATION: /product= "Module 1 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 24638..30820
OTHER INFORMATION: /product= "Module 2 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 30881..35446
OTHER INFORMATION: /product= "Module 3 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 35528..40111
OTHER INFORMATION: /product= "Module 4 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 40190..46318
OTHER INFORMATION: /product= "Module 5 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 46851..47891
OTHER INFORMATION: /product= "SorM"
OTHER INFORMATION: /note= "The protein encoded by the sorM gene is highly
OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the
OTHER INFORMATION: polyketide rapamycin."
US-08-764-233A-1

Query Match      3.8% Score 47.6; DB 2: Length 49377;
Best Local Similarity 44.1%; Pred. No. 0.0089; Indels 3; gaps 2;
Matches 286; Conservative 0; Mismatches 359;

OY 154 GCCATGAGTGTCTCATCCAGGACAGGTGCCGAGAGGGCTTCGGTATCCAGCAGCGGAT 213
    | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41609 GCCCACGACGACTACCTCGCGGATGTGCCTATTGCTTGCGACCAACCAGCGGCCAC 41668
    | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 214 TTCAAAGGCGGACAGCGCCCAACTGGTCGCGACCGCCGACACAGGTGTGGCCGACGCTGAT 273
    | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41669 TTCGAGCACCGCGCGCTCTGTGTAGCCCAACGCGACGAGCTCTCTCCGCGCTCGAC 41728
    | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 274 TTATGCTCAAGGTCCAAGAACCAGTATAGCGGGGATATCGGCCCTTCGACACAGGGCAG 333
    | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41729 TCGCTCGCCACGACMAACCCCGCCCCGACAGACGCTCTCTGGAAGGAAAGCCACGGCC 41788

```

QY 334 ATCTGTACGTTCTTGTGATTTGGCCGCGTCAAGTCTTGACACCGATGCTTGTGAT 393  
 DB 41789 AAGCTGCTCTTCTTCTTCTTGCGCAGAGGCTCGAGTGGGAGAGGAGTGGCCCTGCTG 41848  
 QY 394 TCCGGCAGCAGTCAATTGCTTACAGAGCCGTCACAGCCCGGCGCACTACCCCTG 453  
 DB 41849 CTCGACTCTCGCCCGCTTCTCGGCTCAGCTCAAGCATGAGCGCGGCTGCTCT 41908  
 QY 454 CTGCCCCGAGTGAAGGAGGAGCGCGGCTGCACTGCGCCGCAAGTGGCGCTTACAC -C 511  
 DB 41909 CAGTCGAGTGAAGGAGGAGCGCGGCTGCACTGCGCCGCAAGTGGCGCTTACAC 41968  
 QY 512 TGATGCGAAGCCAGGAGGCGCGGCTGCTGATGAGCGGCGTCCGCGCTCAACCG 571  
 DB 41969 CGGCTGACAGCTGTAAGAGCCCGGCTTGTGCGCTCATGCTCTCGGCGGCGCTG 42028  
 QY 572 CCGAGC-TGCTGTGATCGGCGCGGCGCAGCCGCTACAGAGAGCCGCTGAGCCAC 630  
 DB 42029 CGCTGCTGCGGCTGAGAGCCCGGCGGCTGCTGCGCAGTCAAGCGAGATCGCGCC 42088  
 QY 631 GCGATGCGGCGAGCCGCTTACGCTTACATCAACATCAACATTCGCGCACTCGAC 690  
 DB 42089 GCGTCTGCTGAGGCGGCTTCTCTCTGAGAGAGCGGCGGCTGCGCGCTGCGAGC 42148  
 QY 691 GCGGAGTTCTGCGGCGGAGTCACTGCTACTCATGCGCTTACAGAGTGGAGGTC 750  
 DB 42149 AAGCGCTCAACAGCTGCGCGGCGCAGGCGCATGCGCGCTGAGCTGCGGCTCC 42208  
 QY 751 GTCAAGGTGCGGAGCTGATGAGGCGGCTGCTGAGCGGCGGCGGCGGCGGCGG 798  
 DB 42209 GACCTCAGAGCTACTCTCTCTCTGCGGCGCAGAGCTCTCTCATCGCC 42256

## RESULT 11

US-08-816-105A-2

Sequence 2, Application US/08816105A

Patent No. 5989882

GENERAL INFORMATION:

APPLICANT: Criegee, Anneliese; Dellweg, Hans-Georg; Lenz,  
 APPLICANT: Jurgen; Schröder, Werner; Pape, Hermann;  
 APPLICANT: Goetze, Klaus; Schaper, Beate; Hemker, Michael;  
 APPLICANT: Piepersberg, Wolfgang; Distler, Jürgen;  
 APPLICANT: Strümann, Ansgar  
 TITLE OF INVENTION: PROCESSES FOR PREPARING ACAROTOSYL  
 TITLE OF INVENTION: TRANSFERASE AND FOR USING IT IN THE  
 TITLE OF INVENTION: CONVERSION OF ACARBOSE HOMOLOGUES  
 TITLE OF INVENTION: INTO ACARBOSE, FOR THE PREPARATION  
 TITLE OF INVENTION: OF ACARBOSE HOMOLOGUES  
 NUMBER OF SEQUENCES: 16  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sprung Kramer Schaefer & Briscoe  
 STREET: 660 White Plains Road  
 CITY: Tarrytown  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10591-5144

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4 MB storage

COMPUTER: Apple Macintosh 6500

OPERATING SYSTEM: System 7.5

SOFTWARE: WordPerfect 3.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/816, 105A

FILING DATE: 14-MAR-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE 19625269.5

FILING DATE: 25-JUN-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE 19611252.4

FILING DATE: 22-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Kurt G. Briscoe

REGISTRATION NUMBER: 33,141  
 REFERENCE/DOCKET NUMBER: Bayer 9814-KGB  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (914) 332-1700  
 TELEFAX: (914) 332-1844  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2582 nucleotides  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 US-08-816-105A-2

## Query Match

Best Local Similarity 47.3%; Score 46; DB 4; Length 2582;

Matches 139; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 359 CCGGCTACGCTGCTTGCACCGATGCGTTGTGATTCGCGCAGCAGTCAATTGCTTACG 418  
 DB 1576 CCGGCTGCTGCGGCGGCGGCTTATGCGAGCTGCTGCGGCGGCTTCCGAGC 1635  
 QY 419 AGACGCTGACAGCGCGGCGGCGGCTTATGCGAGCTGCTGCGGCGGCTTCCGAGC 478  
 DB 1636 TGACCGTGAAGAGAGCGGCGGCGGCTTATGCGAGCTGCTGCGGCGGCTTCCGAGC 1695  
 QY 479 GTCGACTGCGGCGGCGGCTTATGCGAGCTTATGCGAGCTTATGCGAGCTTATGCGAG 538  
 DB 1696 GCGAGTGTGCGGCTTATGCGAGCTTATGCGAGCTTATGCGAGCTTATGCGAGCTTATGCGAG 1755  
 QY 539 TGCTGATGCGGCGGCGGCTTATGCGAGCTTATGCGAGCTTATGCGAGCTTATGCGAGCTTATGCGAG 598  
 DB 1756 TCGGCGGCTGCTGATGCGGCGGCGGCTTATGCGAGCTTATGCGAGCTTATGCGAGCTTATGCGAGCTTATGCGAG 1815  
 QY 599 CCGGCGGCTGCTGATGCGGCGGCGGCTTATGCGAGCTTATGCGAGCTTATGCGAGCTTATGCGAGCTTATGCGAG 652  
 DB 1816 CCGGCGGAGACCGTTCGATGCGGCGGAGTCCCCCGGAGCTTATGCGAGCTTATGCGAGCTTATGCGAGCTTATGCGAG 1869

## RESULT 12

US-08-403-852D-3

Sequence 3, Application US/08403852D

Patent No. 5891695

GENERAL INFORMATION:

APPLICANT: Blanc, Veronique  
 APPLICANT: Blanchet, Francis  
 APPLICANT: Crouzet, Joel  
 APPLICANT: Jacques, Nathalie  
 APPLICANT: Lacroix, Patricia  
 APPLICANT: Thibaut, Denis  
 APPLICANT: Zagorec, Monique  
 APPLICANT: Debussche, Laurent  
 APPLICANT: De Crey-Lagard, Valerie  
 TITLE OF INVENTION: Polypeptides Involved In The  
 TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
 TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
 NUMBER OF SEQUENCES: 43  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
 STREET: 1300 I Street, N.W., Suite 700  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20005-3115

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/403, 852D

FILING DATE: 10-MAY-1995

PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA: FR 92/11441
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..833
US-08-403-852D-3

```

```

Query Match      3.6%; Score 45.4; DB 3; Length 833;
Best Local Similarity 48.0%; Pred. No. 0.009;
Matches 130; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```

```

QY 549 CGGGGTCCCGGCGTGAACCGGCGGAGCTGTGATGCGGCGCGGCGGCTA 608
    || || || || || || || || || || || || || || || || || ||
DB 327 CGAGGGCGCGACCGGCACTCATCGAGCTGTCCGGGCTGTGGACAGCTTCA 386
QY 609 CAACGACGCCGCGATCGCCCAAGCGCATGGCGCGACCGCTT 368
    || || || || || || || || || || || || || || || || || ||
DB 387 CGAGCGCTGTGTCACGACCGCGCGGCGGCTGTACTGGGCGCTGCCGCTCC 446
QY 669 CGACAACTTCGGAATCGACGCGGAGTTCTGGGCGCGATCCACTGCTCTCATC 728
    || || || || || || || || || || || || || || || || || ||
DB 447 ACTGACCGACCGGCGAGCGACTTTCGAGTGGCGCGCCCTCAACGTCGCCGCC 506
QY 729 GGCCTACGACTCGAGGCGCGCTCAAAAGCTGCCACTGTGATTTGGGCGCGTCTGT 788
    || || || || || || || || || || || || || || || || || ||
DB 507 GCAGGGCGACCCCTGCTGCTCCGCTACCGGCGCGCCCTTCCGCCGCGGACCT 566
QY 789 GCCAGGCGCCCAAGCGCACCAATTAGTCTCG 819
DB 567 CCTGCTCGAGGAGGCGCGCGACGCGGCTCG 597

```

```

RESULT 13
US-08-403-852D-1
; Sequence 1, Application US/08403852D
; Patent No. 5891695
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanche, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debussche, Laurent
; APPLICANT: De Crey-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved In The
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
; NUMBER OF SEQUENCES: 43
; CODING FOR THESE POLYPEPTIDES AND THEIR USE
; CORRESPONDENCE ADDRESS:

```

```

ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852D
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5392 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
US-08-403-852D-1

```

```

Query Match      3.6%; Score 45.4; DB 3; Length 5392;
Best Local Similarity 48.0%; Pred. No. 0.016;
Matches 130; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```

```

QY 549 CGGGGTCCCGGCGTGAACCGGCGGAGCTGTGATGCGGCGCGGCGGCTA 608
    || || || || || || || || || || || || || || || || || ||
DB 3018 CGAGGGCGCGACCGGCACTCATCGAGCTGTCCGGGCTGTGGACAGCTTCA 3077
QY 609 CAACGACGCCGCGATCGCCAAAGCGCATGGCGCGACCGTTACGTTCTAGACATCAACAT 668
    || || || || || || || || || || || || || || || || || ||
DB 3078 CGAGCGCTGTGTCACGACCGCGCGGCGGCTGTACTGGGCGCTGCCGCTCC 3137
QY 669 CGACAACTTCGGAATCGACGCGGAGTTCTGGGCGCGATCCACTGCTCTCATC 728
    || || || || || || || || || || || || || || || || || ||
DB 3138 ACTGACCGACCGGCGAGCGACTTTCGAGTGGCGCGCCCTCAACGTCGCCGCC 3197
QY 729 GGCCTACGACTCGAGGCGCGCTCAAAAGCTGCCACTGTGATTTGGGCGCGTCTGT 788
    || || || || || || || || || || || || || || || || || ||
DB 3198 GCAGGGCGACCCCTGCTGCTCCGCTACCGGCGCGCCCTTCCGCCGCGGACCT 3257
QY 789 GCCAGGCGCCCAAGCGCACCAATTAGTCTCG 819
DB 3258 CCTGCTCGAGGAGGCGCGACGCGGCTCG 3288

```

```

RESULT 14
US-08-403-852D-6
; Sequence 6, Application US/08403852D
; Patent No. 5891695
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanche, Francis

```

APPLICANT: Crouzet, Joel  
APPLICANT: Jacques, Nathalie  
APPLICANT: Lacroix, Patricia  
APPLICANT: Thibaut, Denis  
APPLICANT: Zagorec, Monique  
APPLICANT: Debussche, Laurent  
APPLICANT: De Grey-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved In The  
Biosynthesis Of Streptogramins, Nucleotide Sequences  
TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pinegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,852D  
FILING DATE: 10-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806, 0054-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1833 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.pristinaespiralis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 103..1689  
US-08-403-852D-6

Query Match 3.4%; Score 42.4; DB 3; Length 1833;  
Best Local Similarity 50.5%; Pred. No. 0.064;  
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

539 TGCTGATGGCGGGGTCGCCGCGTCAACCGCGCGATGCTGTGATCGCGCGCGCA 598  
413 TGAATGCTCGAGACTGCTCTTCTGCGCGCGCGTCTGCGCGCGCGCGCGCGCA 472  
599 CGCGCGGCTCAACG 658  
473 CGCA 532  
659 ACATCAACATGCAACACTTGGGAACTGACGCGCGCGCGCGCGCGCGCGCGCG 718  
533 CGACCGCTCG 592  
719 GCTACTCATCG 742

Db 593 TGTGGGAGCG 616  
RESULT 15  
US-08-173-508-3  
Sequence 3, Application US/08173508  
Patent No. 5616485  
GENERAL INFORMATION:  
APPLICANT: Bartfield, Daniel  
APPLICANT: Butler, Michael J.  
APPLICANT: Hadary, Dany  
APPLICANT: Jenish, David  
APPLICANT: Krieger, Timothy  
TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED  
STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND  
TITLE OF INVENTION: POLYPEPTIDES  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/173,508  
FILING DATE: 23-DEC-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 18740/125/CACO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202 672 5300  
TELEFAX: 202 672 5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2185 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 531..2069  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 531..902  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 903..2069  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 531..533  
OTHER INFORMATION: /note="Met at position -124  
represents fMet"  
US-08-173-508-3

Query Match 3.4%; Score 42.4; DB 1; Length 2185;  
Best Local Similarity 47.4%; Pred. No. 0.067;  
Matches 127; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

531 CGCGGCTGCTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 590  
Db 1250 CGCGGCTGCG 1309  
591 CGCGGCG 650

Fri Jun 23 09:31:07 2000

us-09-362-485-2.mli

Page 12

Db	1310	CCCCGAGCGCTTCTTCTACACCGAAGCCGCTGCTCTCGCGCTTCAATGTCGCGCGCCGAGCA	1369
QY	651	GGTTCTAGACATCAACATCGACAAACTTCGCGCACTCGACGCCGAGTTCTCGGCGCGGAT	710
Db	1370	CGGGGTGACGTGACCAACAACAGCTATTACACGACCCCGTGGTACTTCACTGGAAGA	1429
QY	711	CCACACTCGCTACTCATCGGCTACGAGGTCGAGGGTGGCGTCAAACTGCCGACCTGGT	770
Db	1430	CGACCCCGACCAAGAAAGCGCTGCTCGTAGGGCGGCTCTCGCGGCGCTCCGGTAGCGGAGAA	1489
QY	771	GATTGGGGCCGCTCCTGGTGGCAGGCGCC	798
Db	1490	GAGGGGCGCGGTCAACGTCGCGCGGGCC	1517

Search completed: June 22, 2000, 15:01:15  
Job time: 17553 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:07:07 ; Search time 5541.94 Seconds  
(without alignments)  
910.559 Million cell updates/sec

Title: us-09-362-485-2

Perfect score: 1245  
Sequence: 1 ATCTGCAGATTATTCGAC.....AAGCGATGATCGCGCCG 1245

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 202611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: gb\_est33:\*  
53: gb\_est34:\*  
54: gb\_est35:\*  
55: gb\_est36:\*  
56: gb\_est37:\*  
57: gb\_est38:\*  
58: gb\_est39:\*  
59: gb\_est40:\*  
60: gb\_est41:\*  
61: gb\_est42:\*  
62: gb\_est43:\*  
63: gb\_est44:\*  
64: gb\_est45:\*  
65: gb\_est46:\*  
66: gb\_est47:\*  
67: gb\_est48:\*  
68: gb\_est49:\*  
69: gb\_est50:\*  
70: gb\_est51:\*  
71: gb\_est52:\*  
72: gb\_est53:\*  
73: gb\_est54:\*  
74: gb\_est55:\*  
75: gb\_est56:\*  
76: gb\_est57:\*  
77: gb\_est58:\*  
78: gb\_est59:\*  
79: gb\_est60:\*  
80: gb\_est61:\*  
81: gb\_est62:\*  
82: gb\_est63:\*  
83: gb\_est64:\*  
84: gb\_est65:\*  
85: gb\_est66:\*  
86: gb\_est67:\*  
87: gb\_est68:\*  
88: gb\_est69:\*  
89: gb\_est70:\*  
90: gb\_est71:\*  
91: gb\_est72:\*  
92: gb\_est73:\*  
93: gb\_est74:\*  
94: gb\_est75:\*  
95: gb\_est76:\*  
96: gb\_est77:\*  
97: gb\_est78:\*  
98: gb\_est79:\*  
99: gb\_est80:\*  
100: gb\_est81:\*  
101: gb\_est82:\*  
102: gb\_est83:\*  
103: gb\_est84:\*  
104: gb\_est85:\*  
105: gb\_est86:\*  
106: gb\_est87:\*  
107: gb\_est88:\*  
108: gb\_est89:\*  
109: gb\_est90:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

%  
Query

SUMMARIES





## FEATURES

## 1.844

BASE COUNT	261 a	112 c	92 g	35 t	344 others
ORIGIN					

[illegible]

```
D6      |::|:|:|.:. . . . .  
768 CCMVCTTTHTTYTCMYAMMBRYAMMMMAMMTTMCVMMAVAVVYAAAMAMB 709
```

382

Db 708 NNVYCSBSCNBTITYCNYTCVSSSBSSBTTCWTHHTNSNTYTMVTVTTHHTYTNTC 649

203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1

**THE UNIVERSITY OF CHICAGO**

Db 588 CGCYSBVCSSBSTBSSVSGBSSTSSGTCCKCBSSSSSBBBCSSBSSCCCYCSSYCTYCY 529

0Y 30Z GURIALCCLIGMIGCGAACCCMHBBBCCCCCCCCCCCCCCCCCC  
.:|.: :|.: :|.: :|.: :|.: :|.: :|.: :|.: :|.: :|.

QY 562 GTCGAACCGCGCGACGTCGTGTGATCGGCGCGCGCACCGCGCGCTACACGCGAGCCGC 621

2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526

108 WYVSESSCCCTC 392

**RESULT 5**

LOCUS Drosophila melanogaster genome survey sequence T7 en' of BAC #  
 DEFINITION

fly), genomic survey sequence.  
AT055630

SOURCE fruit fly.

**Eukaryota:** Eudicotyledons; Monocotyledons; Gymnosperms; Angiosperms; Pteridophytes; Bryophytes; Fungi; Metazoa; Mollusca; Annelida; Nematoda; Arthropoda; Chordata; Vertebrates.

**AUTHORS** Genoscope.

BP 191 91006 EVRY cedex - FRANCE (E-mail : [segretef@genoscope.cns.fr](mailto:segretef@genoscope.cns.fr))

## FEATURES

## 1. .910

ORIGIN

Query Match 3.98: Score 49: DB 82: Length 91

[illegible]

Db 713 SRVMSAGSSASSSSASSSSASSSSASMVSCRSASVSVASASAAAS

[illegible]

519 AACCCAAGGGGCCCGCTC ;CTGATGGCGGGGTGCCCGGGC; :GAACCG

579 CGTGGTAGATCCGCCGCGCAGCCGCGCTACAAACGAGCCCGCATCGGC 62

LOCUS	CNS010BJ	845 bp	DNA	GSS
CNS010BJ/C				

BACN04N13 of DrosBAC library from *Drosophila melanogaster*

**KEYWORDS** GSS.

ORGANISM      PROSOPHILID      MELALLOGASTRI  
Eukaryota:      Metazoa:      Arthropoda:      Tracheata:      Hexapoda

REFERENCE 1 (bases 1 to 846)

Document submitted (23 Oct 1999) Confidential  
 BP 191 91005 EVRY cedex - FRANCE (E-mail) : sequef

collaboration with the European Drosophila genome

project grant. The DNA was prepared from embryos

**FEATURES**

**source** pBelOBAC11.

**Location/Qualifiers**

1..846  
/organism="Drosophila melanogaster"  
/plasmid="pBelOBAC11"  
/db\_xref="taxon:7227"  
/clone\_1fb="DrosBAC"  
/clone="BACN04N13"  
/note="end : 77"

**BASE COUNT** 300 a 47 c 40 g 175 t 284 others

**ORIGIN**

Query Match 3.7% Score 46.6; DB 82; Length 846;  
Best Local Similarity 13.0%; Pred. No. 6.7;  
Matches 33; Conservative 127; Mismatches 93; Indels 0; Gaps 0;

Dy 369 TCGTTCACCCGATGGCTTGATGCCGAGATCCGCACACGATCAATTGCTCAGACAGCCTCCA 428  
|||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::  
Db 842 TACTSSSTSTSCSSSSSBTTSTTTSTTSASSSSSSTSSSTSSSTSSSTSSSTSSBTBT 783  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Dy 429 GACCCCGCAGCGCACATCCCTCGTTTCCCCGATGACCAGAAGTGGCCGGGTGATCCC 488  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Dy 782 SSSSTTTSTTTSSSTTTTACCATCTGATCGAACCAGGAGG.CCGCGGTGCTGATGGG 723  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::

Dy 722 SBSTGSTSTSBRTBTBTBSSSBSRSSSSSTSTSTBTBSSSSSSBRTBSTSSSTTSS 663  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::

Dy 549 CGGGGTGCGCCGCGTTCGAACCGCGCAGCTCGTGTGATCGCGCGGACCGCGGCTTA 608  
::::::::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::

Dy 662 SSSSBSTSTSSBSSTCTTSTTSSSSCSSSSSTSTSSBSBSBCBTSSCSSSSSCCT 603  
::::::::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::

Dy 609 CAAACGACGCGCG 621  
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::

Dy 602 CGCCSCGCCCCCC 590  
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::

**RESULT** 7  
**CNS0181E/c**

**LOCUS** CNS0181E/c

**DEFINITION** Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37H05 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

**ACCESSION** AL108764

**VERSION** AL108764.1 GI:5629068

**KEYWORDS** GSS.

**SOURCE** fruit fly.  
**ORGANISM** Drosophila melanogaster  
Euarthropoda, Metazoa, Arthropoda, Tracheata, Hexapoda, Insecta; Pterygota, Neoptera, Endopterygota; Diptera, Brachycera; Muscormorpha; Tephrozoidea; Drosophilidae; Drosophila.

**REFERENCE** 1 (bases 1 to 1101)  
**AUTHORS** Genoscope.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr)

**COMMENT** Web : www.genoscope.cns.fr  
determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billand at CPBH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

```
FEATURES
source
location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/assembly="pBel0BAC11"
/db_xref="taxon:7227"
/clone_1fb="DrosBAC"
/clone="BACN37H05"
```

BASE COUNT	61 a	215 c	301 g	235 t	289 others
ORIGIN	/note="En.: SP6"				
Query Match	3.7%; Score 46.6; DB 83; Length 1101;				
Best Local Similarity	40.2%; Pred. No. 6.9;				
Matches 101; Conservative	40; Mismatches 109; Indels 1; Gaps 1				
OY	395	CCGGACACCAGTCATTTGCTACTAGAGACCGTCTCAACACGCCGACGGCAGTACCCCTGC	454		
Db	308	CCCCAIVVVVGMMAAACSCCCCAVAASCSGCCSCCCSCCCSCCCSCCSCSACS	249		
OY	455	TTGACCCCATGAGACCAATCGCGGTCTGCAGTCCGCGCGCCAGGTGTGGCGTTACCACTTA	514		
Db	248	AAGMGSMISGVGGACAAGAAG	GAGACCCCGCCSCACSCCSCSSCGCSCSSCSCSSSS	189	
OY	515	TGCGAACCCAAAGGGGGCGCGGTGTGCTGATGAGGGGGTGCCCGGCGTGAACCGCGG	574		
Db	188	CGCCACCCCSSAGSGGGCCCGCCGCCACCCCGCCSGAGCAVAGCCGCCAC	-CMNVTGGGGG	130	
OY	575	ACGTGTGGTGTATCGGGCGCCGCGCACCGCGGTATACAGCAGCAGCCCGCATCGCCACGCA	634		
Db	129	ACGCAAGGGAGAGAGCGCGGNAAGCCCGCNCACCCCGCCCGCCAGCGCGGCGCA	70		
OY	635	TGGGCGCGAC	645		
Db	69	AGCGCCCAACC	59		
RESULT	8				
CNS006XK					
LOCUS	CNS006XK 935 bp DNA GSS 03-JUN-1999				
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC #				
VERSION	BACR1N09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
KEYWORDS	AL066051.1 GI:4945019 GSS.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscumorphia; Ephydroidea; Drosophilidae; Drosophila. Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.				
COMMENT					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
FEATURES					
Source	Location/Qualifiers 1..935 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPCI-98" /clone="BACR1N09" /note="End : T7"				
BASE COUNT	257 a	170 c	162 g	96 t	250 others







Qy	546	GGGGGGGGTGC	CCCCCGTCG	AAACCGCGACAGCTCTGGATCGGGCGGACCGCGG	605
		.....	.....	.....	
Db	875	SSSSSVS	AAAAVS	SSSSSASMSAMAAAAVA	VSSVS
		.....	.....	.....	.....
Qy	606	CTACACG	CAGCCCGCATCGCC	ACGCGATGGCGCGACCGTTACGGTTACGACATCAA	665
		.....	.....	.....	.....
Db	935	SVMA	ASSSVSSSSSVSTSS	SVSSVSAVSMASVSSSSSSSVSSSVVA	AAASAA
		.....	.....	.....	.....
Qy	666	CATCGCA	AAACTTCGGCAACTCGACG	CCGAGTTCGGCGCGGATCCGACATCGGTATC	725
		.....	.....	.....	.....
Db	995	AAAA	AAAASSSSSAAVA	VSSSSSSSSSSSSSSSSSVSSSVSSSCSVSVSSSS	1054
		.....	.....	.....	.....
Qy	726	ATCGGCT	TCGAGCTCGA	743	
		.....	.....	.....	.....
Db	1055	VSVAVAS	ASASASVSSVA	1072	

RESULT	13				
LOCUS	A1712257				
DEFINITION	A1712257	315 bp	mRNA	EST	02-FEB-2000
	6050690909.x1	605	- Endosperm	CDNA library from	Schmidt lab Zea mays
ACCESSION	A1712257				
VERSION	A1712257.1	GI:5006195			
KEYWORDS	EST.				
SOURCE	Zea mays.				
ORGANISM	Zea mays				

REFERENCE	1 (bases 1 to 315)
AUTHORS	Maiboh, V.
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL	Unpublished (1999)
COMMENT	On Jun 22, 1998 this sequence version replaced gi:3246668.

JOURNAL  
COMMENT

Unpublished (1999)  
On Jun 22, 1998 this sequence version replaced gi:324668.  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 605069 row: D column: 09.

FEATURES	Location/Qualifiers
source	1. .315
	/organism="Zea mays"
	/cultivar="Oh1043"
	/db_xref="taxon:4577"
	/clone_1ib="605 - Endosperm cDNA library from Schmidt lab
	/tissue_type="nucellar, embryo, and endosperm"
	/dev_stage="10-14 days post-pollination"
	/lab_host="DH5(alpha)"
	/note="Organ: kernel; Vector: pAD-GAL4-2'; Site_1: EcoRI; Site_2: XhoI; kernel endosperm cDNA library from Schmidt lab"
BASE COUNT	61 a 105 c 106 g 42 t 1 others
ORIGIN	

Query Match	3.6%	Score 45	DB 51	Length 315
Best Local Similarity	50.7%	Pred. No. 12		
Matches 108	Conservative	0	Mismatches 105	Indels 0
			Gaps	
QY 397	GGGACACGTCATTGCTCCACGAGACCGGCCAGACGGCGGCACTACCCCTGTT			
Db 58	GGCATGGCGTCTTCATCCAAAGACCAAGCTGGGTGGCGGACCAACTTTTAAACACCCC			
QY 457	GGCCCGATGGCGAAGTCCGCTGGACACGGCCGCCAGTTGGCGCTTTACACCTGATG			
Db 118	GGCCTTAGCAGCCAGGCGCGGCTCAGCGGATTAAGGGCGGGCGGGCGTGGGAGACACCG			

Qy 517 CGAACCCAAAGGGGGCCGGGGGTCTATGGCCGGGTGCCCGGCGTGCAGAACGCGCGAC 576

Db 178 CGTGGCGACACGGGGCCGGCGCTCTCTTTATCTTCTTCTTCTGCTGGCCGCGAGGC 237

Qy 577 GTCCGCTGATCGGCGCCGACGACCGCGCGCTAC 609

Db 238 GTTGGCGGTGGCCGGCCAGCGCGTCCGCTCC 270

RESULT 14				
CNS010RJ				
LOCUS	846 bp	DNA		
DEFINITION	CNS010RJ	GSS	26-JUL-1999	
	Drosophila melanogaster genome survey sequence T7 end of BAC			

ACCESSION	AL099337
VERSION	AL099337.1
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster

REFERENCE 1 (bases 1 to 846)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Séquençage  
Submitted (23-JUL-1999) Genoscope - Centre National de Séquençage  
Submitted (23-JUL-1999) Genoscope - Centre National de Séquençage

COMMENT

Determination of this BAC-end sequence was carried out as part of collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CEH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

FEATURES	source	location/Qualifiers
	1. 846	
	/organism="Drosophila	melanogaster"
	/plasmid="pbeloBAC11"	
	/db_xref="taxon:7227"	
	/clone_1ib="DrosBAC"	
	/clone_1BACN04N13"	
	/note="and : T7"	
BASE COUNT	300 a	47 c 40 g 175 t 284 others
ORIGIN		

Query Match	3.6%	Score 45	DB 82	Length 846
Best Local Similarity	16.2%	Pred No. 14		
Matches 38	Conservative 10%	Mismatches 88	Indels 0	Gaps 0

512 TGATGCGAACCCAGGGGCCCGGTGTGCTGATGGCGGGGTGCCCGGCTCGAACCG 571

Db 574 KGAGCGRARSSAGAGGGGGSSGGCCGAGCGSSSSSSGSSAVGSSSSSVSSASSASSS 633

QY 572 CCGACGTCGTGGTATCGGGCCGGCACC GGCGCTACAACGCAGCCCGCATCGCCAACG 631

634 SSGSSSSAASAAGACSSVSSASSAVSSSSSSAASAASSASASVSSSSSVAVAASSASS 693

[illegible]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

DB 094 SSISVSSSSAAVAVVSASSACASVSSSVSCAASSASASSISISSSSSAAASSSSSSSS / 2

QY 692 CCGAGTCTCGGGCCGGATCCACACTCGCTACTCATCGGCTACGAGCTCGAGGG /46

Db 754 SSASGASSSSAASAAASSSSAAASAVAVASASSSSASASSSSSSASSSSSSS 808

Journal of  
Management  
Education  
33(1)

RESUL 13  
CNS0072Q

LOCUS	CNS00720	932 bp	DNA	GSS	03-JUN-1999
-------	----------	--------	-----	-----	-------------

Downloaded from <http://ajphaphysocpharm.sagepub.com/> at 11:01 11 November 2014



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:50:58 ; Search time 8627.09 Second/  
(Without alignments)  
-139.259 Million cell updates/sec

Title: US-09-362-485-3  
Perfect score: 1235  
Sequence: 1 ATCTTCAGATTAATCGAAC.....ANGAAGCGATGATCGNC 1235

Scoring table: IDENTITY\_NTC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 segs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765534

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_da1:\*  
2: gb\_da2:\*  
3: gb\_om:\*  
4: gb\_om:\*  
5: gb\_om:\*  
6: gb\_ph:\*  
7: gb\_ph:\*  
8: gb\_ph:\*  
9: gb\_ph:\*  
10: gb\_ph:\*  
11: gb\_ph:\*  
12: gb\_ph:\*  
13: gb\_ph:\*  
14: gb\_ph:\*  
15: gb\_ph:\*  
16: gb\_ph:\*  
17: gb\_ph:\*  
18: gb\_ph:\*  
19: gb\_ph:\*  
20: gb\_ph:\*  
21: gb\_ph:\*  
22: gb\_ph:\*  
23: gb\_ph:\*  
24: gb\_ph:\*  
25: gb\_ph:\*  
26: gb\_ph:\*  
27: gb\_ph:\*  
28: gb\_ph:\*  
29: gb\_ph:\*  
30: gb\_ph:\*  
31: gb\_ph:\*  
32: gb\_ph:\*  
33: gb\_ph:\*  
34: gb\_ph:\*  
35: gb\_ph:\*  
36: gb\_ph:\*  
37: gb\_ph:\*  
38: gb\_ph:\*  
39: gb\_ph:\*  
40: gb\_ph:\*  
41: gb\_ph:\*  
42: gb\_ph:\*  
43: gb\_ph:\*  
44: gb\_ph:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1232	99.8	1235	5	A87606 Sequence 3
2	1232	99.8	1235	5	A89746 Sequence 3
3	1232	99.8	1236	5	A87613 Sequence 10
4	1232	99.8	1236	5	A89753 Sequence 10
5	1230.4	99.6	1237	5	A87607 Sequence 4
6	1230.4	99.6	1237	5	A89747 Sequence 4
7	1230.4	99.6	56414	1	MTW002
8	1224.4	99.1	1228	5	A87608 Sequence 5
9	1224.4	99.1	1228	5	A89748 Sequence 5
10	1219	98.7	1235	5	A87609 Sequence 6
11	1219	98.7	1235	5	A89749 Sequence 6
12	1218.4	98.7	1235	5	A87611 Sequence 8
13	1218.4	98.7	1235	5	A89751 Sequence 8
14	1216	98.5	1245	5	A87605 Sequence 2
15	1216	98.5	1245	5	A89745 Sequence 2
16	1216	98.5	1260	5	A87604 Sequence 1
17	1216	98.5	1260	5	A89744 Sequence 1
18	1216	98.5	2584	1	MTALADH
19	1213.4	98.3	1229	5	A87610 Sequence 7
20	1213.4	98.3	1229	5	A89750 Sequence 7
21	1207	97.7	1209	5	A87612 Sequence 9
22	1201	97.2	1208	5	A89752 Sequence 9
23	1194	96.7	1194	2	MTU92472
24	698	56.5	720	5	A89759 Sequence 16
25	538.4	43.6	40745	1	SC151
26	467.6	37.9	682	5	A89758 Sequence 15
27	467.6	37.9	682	5	A89762 Sequence 19
28	377	30.5	8046	2	AF049107
29	351.2	28.4	11514	2	AE002028
30	351.2	26.8	2780	1	AE002028
31	330.8	26.8	1585	2	AE002028
32	312.8	25.3	1736	1	AE002028
33	293	23.7	1477	2	AE002028
34	288.4	23.4	2113	2	AE002028
35	277.2	22.4	1125	5	AE002028
36	258	20.9	13740	1	AE002028
37	243.2	19.7	1375	1	AE002028
38	243	19.7	1776	1	AE002028
39	235.6	19.1	1709	1	AE002028
40	235.6	19.1	12316	1	AE002028
41	235.6	19.1	217420	1	AE002028
42	218.2	17.7	1714	2	AE002028
43	107.2	8.7	10652	2	AE002028
44	99.2	8.0	10512	2	AE002028
45	74.2	6.0	3160	1	AE002028

## ALIGNMENTS

```

RESULT 1
AB7606 1235 bp DNA PAT 22-JAN-2000
LOCUS Sequence 3 from Patent W09836089.
DEFINITION A87606
ACCESSION A87606
VERSION A87606.1 GI:6736246
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1235)
AUTHORS Flohe, L. and Singh, M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
source 1..1235
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 394 c 382 g 220 t 3 others
ORIGIN
Query Match 99.8%; Score 1232; DB 5; Length 1235;
Best Local Similarity 100.0%; Pred. No. 3.3e-169;
Matches 1235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTTGAGATTATGCAACTTTCTTCACTGAGCTAGCATGATGAGAGGGTATC 60
DB 1 ATCTTGAGATTATGCAACTTTCTTCACTGAGCTAGCATGATGAGAGGGTATC 60
QY 61 ATGCGCTGCTGATTCGACGAGCAACAAAAACAAGCAATCCGGTGGCCATACCCG 120
DB 61 ATGCGCTGCTGATTCGACGAGCAACAAAAACAAGCAATCCGGTGGCCATACCCG 120
QY 61 ATGCGCTGCTGATTCGACGAGCAACAAAAACAAGCAATCCGGTGGCCATACCCG 120
DB 61 ATGCGCTGCTGATTCGACGAGCAACAAAAACAAGCAATCCGGTGGCCATACCCG 120
QY 121 GCCGCGTGGGAGCACTAACCCGCTGAGCATGAGTGTCTCATTCAGCAGAGTCCGGA 180
DB 121 GCCGCGTGGGAGCACTAACCCGCTGAGCATGAGTGTCTCATTCAGCAGAGTCCGGA 180
QY 181 GAGGCTCGGCTATACCGAGCGGGATTTCAAGCGCGGCGGCACTGTTGGGCAAC 240
DB 181 GAGGCTCGGCTATACCGAGCGGGATTTCAAGCGCGGCGGCACTGTTGGGCAAC 240
QY 241 GCCGACGAGGTGGGGCGAGCTGATTTATGTCAGAGTCAAGCAAGCAAGCGGCG 300
DB 241 GCCGACGAGGTGGGGCGAGCTGATTTATGTCAGAGTCAAGCAAGCAAGCGGCG 300
QY 301 GAATACGCGCGCTGCGACAGCGGAGATCTTGTTCAGCTTCTTCATTGGCGCTCA 360
DB 301 GAATACGCGCGCTGCGACAGCGGAGATCTTGTTCAGCTTCTTCATTGGCGCTCA 360
QY 361 CGTGTGACCGATGCTGTTGATTCGCGGACACGTCATGAGTCCAGCAACGCTC 420
DB 361 CGTGTGACCGATGCTGTTGATTCGCGGACACGTCATGAGTCCAGCAACGCTC 420
QY 361 CGTGTGACCGATGCTGTTGATTCGCGGACACGTCATGAGTCCAGCAACGCTC 420
DB 361 CGTGTGACCGATGCTGTTGATTCGCGGACACGTCATGAGTCCAGCAACGCTC 420
QY 421 CAGACGCGGAGCGGCGCACTACCCCTGTTGCCCCGATGACGAAATGCGCGGTGACTC 480
DB 421 CAGACGCGGAGCGGCGCACTACCCCTGTTGCCCCGATGACGAAATGCGCGGTGACTC 480
QY 481 GCCGCCAGAGTTGGGCTTACACCTGATGCGAACCAAGGGGGCGGCTGTGTGATG 540
DB 481 GCCGCCAGAGTTGGGCTTACACCTGATGCGAACCAAGGGGGCGGCTGTGTGATG 540
QY 481 GCCGCCAGAGTTGGGCTTACACCTGATGCGAACCAAGGGGGCGGCTGTGTGATG 540
DB 481 GCCGCCAGAGTTGGGCTTACACCTGATGCGAACCAAGGGGGCGGCTGTGTGATG 540
QY 541 GCGGGGGTGGCGGCGGCGAACCGGCGAGCTGCTGTATCGCGCGCGGCAACCGCGG 600
DB 541 GCGGGGGTGGCGGCGGCGAACCGGCGAGCTGCTGTATCGCGCGCGGCAACCGCGG 600
QY 541 GCGGGGGTGGCGGCGGCGAACCGGCGAGCTGCTGTATCGCGCGCGGCAACCGCGG 600
DB 541 GCGGGGGTGGCGGCGGCGAACCGGCGAGCTGCTGTATCGCGCGCGGCAACCGCGG 600
QY 601 TACAAGCAGCCGCGCATGCGCAAGCGGCGAGCGGCGCATGAGTTAGCATCAAC 650
DB 601 TACAAGCAGCCGCGCATGCGCAAGCGGCGAGCGGCGCATGAGTTAGCATCAAC 650
QY 601 TACAAGCAGCCGCGCATGCGCAAGCGGCGAGCGGCGCATGAGTTAGCATCAAC 650
DB 601 TACAAGCAGCCGCGCATGCGCAAGCGGCGAGCGGCGCATGAGTTAGCATCAAC 650
QY 661 ATGACAAACTTGGGCACTGACGCGGAGTGTGGGGCGGATTCACACTGCTACTCA 720

```

```

DB 661 ATGACAAACTTGGGCACTGACGCGGAGTGTGGGGCGGATTCACACTGCTACTCA 720
QY 721 TCGGCGTACGAGCTGAGAGGTGGCGGTCAAAAGTGCAGCGCTGATGAGGCGCTCTG 780
DB 721 TCGGCGTACGAGCTGAGAGGTGGCGGTCAAAAGTGCAGCGCTGATGAGGCGCTCTG 780
QY 781 GTGCGAGGCGGCGCAAGGCAACCCCAATTAGTCTGCAATTCTGCTGCGCATATGAACCA 840
DB 781 GTGCGAGGCGGCGCAAGGCAACCCCAATTAGTCTGCAATTCTGCTGCGCATATGAACCA 840
QY 841 GGTGCGTACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
DB 841 GGTGCGTACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
QY 901 ACCACCTAGCAACCCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 960
DB 901 ACCACCTAGCAACCCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 960
QY 961 ATGCGCGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
DB 961 ATGCGCGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
QY 1021 GTGCTGAGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1080
DB 1021 GTGCTGAGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1080
QY 1081 AAGGCTTTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
DB 1081 AAGGCTTTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
QY 1141 GTGCGCTTACCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1200
DB 1141 GTGCGCTTACCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1200
QY 1201 CACACNTCGGAGTAAAGC. GCGATGATGTCGNC 1235
DB 1201 CACACNTCGGAGTAAAGCAGATGATGTCGNC 1235

RESULT 2
AB9746 1235 bp DNA PAT 22-JAN-2000
LOCUS Sequence 3 from Patent W09832862.
DEFINITION A89746
ACCESSION A89746
VERSION A89746.1 GI:6738280
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1235)
AUTHORS Flohe, L. and Singh, M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
source 1..1235
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 394 c 382 g 220 t 3 others
ORIGIN
Query Match 99.8%; Score 1232; DB 5; Length 1235;
Best Local Similarity 100.0%; Pred. No. 3.3e-169;
Matches 1235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTTGAGATTATGCAACTTTCTTCACTGAGCTAGCATGATGAGAGGGTATC 60
DB 1 ATCTTGAGATTATGCAACTTTCTTCACTGAGCTAGCATGATGAGAGGGTATC 60
QY 61 ATGCGCTGCTGATTCGACGAGCAACAAAAACAAGCAATCCGGTGGCCATACCCG 120

```

Db	61	ATGGCGTCGGTATTCGGACCGAGAACCAAAAACAAAGCAATTCGCGGCTGGCCATCACCCG	120
Oy	121	GCCGGCGTCGGGAACTAACCCCTCGTGGCCATAGAGTCTCATCCAGCAGTGTCCGGA	180
Db	121	GCCGGCGTCGGGAACTAACCCCTCGTGGCCATAGAGTCTCATCCAGCAGTGTCCGGA	180
Oy	181	GAGGCGTCGGCTATCACCGACCGGAAATTCAGAGCGGACGGCCCACTGGTCGGCAC	240
Db	181	GAGGCGTCGGCTATCACCGACCGGAAATTCAGAGCGGACGGCCCACTGGTCGGCAC	240
Oy	241	GCCGACAGGTGTGGGCGCCGACGCTGATTTATTTGCTCAAGTCAAAAGAACCGATAGCGCG	300
Db	241	GCCGACAGGTGTGGGCGCCGACGCTGATTTATTTGCTCAAGTCAAAAGAACCGATAGCGCG	300
Oy	301	GAATACGGCCCCCGGACACAGCGGAGATCTTGTACGTTCTGTGATTTGGCCGCGTCA	360
Db	301	GAATACGGCCCCCGGACACAGCGGAGATCTTGTACGTTCTGTGATTTGGCCGCGTCA	360
Oy	361	CGTGTTCACACGATGGCTGTGGATTGGATTCGGCACACGCTCAATTTGCTTACAGACGCTC	420
Db	361	CGTGTTCACACGATGGCTGTGGATTGGATTCGGCACACGCTCAATTTGCTTACAGACGCTC	420
Oy	421	CAGACCGCCACAGCGCCGACATACCCCTGCTTGGCCCCGATAGAGAAATGCCGCTGACTC	480
Db	421	CAGACCGCCACAGCGCCGACATACCCCTGCTTGGCCCCGATAGAGAAATGCCGCTGACTC	480
Oy	481	GCCGCGCAGTTTGGCGCTTACACACTATGCGAAACCAAGGGGCGCGGCTGTGCTGATG	540
Db	481	GCCGCGCAGTTTGGCGCTTACACACTATGCGAAACCAAGGGGCGCGGCTGTGCTGATG	540
Oy	541	GCGCGGGTCCCGGCGTGAACCGCGCCGACGTCGTGTATTCGGCGCGGACCGCGCGC	600
Db	541	GCGCGGGTCCCGGCGTGAACCGCGCCGACGTCGTGTATTCGGCGCGGACCGCGCGC	600
Oy	601	TACAAACGAGCGCGCATCGGCAACGAGATGGGGGACGGTTACGGTTCTATACATCAAC	660
Db	601	TACAAACGAGCGCGCATCGGCAACGAGATGGGGGACGGTTACGGTTCTATACATCAAC	660
Oy	661	ATCGACAACCTTCGGGCAACTGACAGCCGATTTGCGGCGCGGATCGACACTGCTACTCA	720
Db	661	ATCGACAACCTTCGGGCAACTGACAGCCGATTTGCGGCGCGGATCGACACTGCTACTCA	720
Oy	721	TGCGGCTACGAGCTCGAGGGTGCCTGTAAACGTGCCGACTGGTATTTGGGGCGTCTG	780
Db	721	TGCGGCTACGAGCTCGAGGGTGCCTGTAAACGTGCCGACTGGTATTTGGGGCGTCTG	780
Oy	781	GTCGACGAGCGCCAGAGGCAACCAATTGTCATTAATCTTGAGCGCATTTGAACCA	840
Db	781	GTCGACGAGCGCCAGAGGCAACCAATTGTCATTAATCTTGAGCGCATTTGAACCA	840
Oy	841	GGTGGGTACTGTGGATATATGACCATGACACGAGGCGGCTGTTTCGAAGGCTCACAGCG	900
Db	841	GGTGGGTACTGTGGATATATGACCATGACACGAGGCGGCTGTTTCGAAGGCTCACAGCG	900
Oy	901	ACCACTACGACACACCGGAGCTTCGCCGTCACAGAACGCTGTTTACGTGCGTGGCGAC	960
Db	901	ACCACTACGACACACCGGAGCTTCGCCGTCACAGAACGCTGTTTACGTGCGTGGCGAC	960
Oy	961	ATGCCCGGCTGGTGCAGAAACGTCACCTACGCGCTGACCAACGCGACGATGCCGAT	1020
Db	961	ATGCCCGGCTGGTGCAGAAACGTCACCTACGCGCTGACCAACGCGACGATGCCGAT	1020
Oy	1021	GTCGTCAGAGCTTGGCGACGAAATGGTGTGGGCGGCGGCTGCGCGGTCAATTCGGGCACTATAC	1080
Db	1021	GTCGTCAGAGCTTGGCGACGAAATGGTGTGGGCGGCGGCTGCGCGGTCAATTCGGGCACTATAC	1080
Oy	1081	AAAGGCTTTTCGAGCGACGAAGAGGGGCTTACTGTCCGAACGGGTGGCCACCGACTGGGG	1140
Db	1081	AAAGGCTTTTCGAGCGACGAAGAGGGGCTTACTGTCCGAACGGGTGGCCACCGACTGGGG	1140
Oy	1141	GTCGCGTTCACCGAGCCCGGACGGTGTGCTGCTGACTCTCGGCGCGCTGTATCCGCGAG	1200
Db	1141	GTCGCGTTCACCGAGCCCGGACGGTGTGCTGCTGACTCTCGGCGCGCTGTATCCGCGAG	1200

OY	1201	CACACNTGGGAGTAANGAAGCAATGATGTGTGC	1235
Dd	1201	CACACNTCGGAGTAGTANGAAGCATGTGATGC	1235
RESULT	3		
LOCUS	A87613	1236 bp	DNA
DEFINITION	Sequence 10 from Patent WO9836089.		PAT
ACCESSION	A87613		22-JAN-2000
VERSION	A87613.1	GI:6736253	
KEYWORDS	.		
SOURCE	unidentified.		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 1236)		
AUTHORS	Flohe, L. and Singh, M.		
TITLE	TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE		
JOURNAL	Patent: WO 9836089-A 20-AUG-1998;		
	FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)		
FEATURES	Location/Qualifiers		
source	1..1236		
	/organism="unidentified"		
	/db_xref="taxon:32644"		
BASE COUNT	236 a 395 c 385 g 220 t		
ORIGIN			
Query Match	99.8%; Score 1232; DB 5; Length 1236;		
Best Local Similarity	99.8%; Pred. No. 3.3e-169;		
Matches 1232:	Conservative 0; Mismatches 3; Indels 0; Gaps 0.		
OY	1	ATCTTGCAATTAAATCGAATTCTTCACACTGGAAGCGTACGATCGAGAGGGTAATC	60
Dd	1	ATCTTGCAATTAAATCGAATTCTTCACACTGGAAGCGTACGATCGAGAGGGTAATC	60
OY	61	ATGGCGGTGGGATTCCGCCGAGACCAAACAACGATTCGGGGTGGCCATCACCCCCG	120
Dd	61	ATGGCGGTGGGATTTCGCCGAGACCAAACAACGATTCGGGGTGGCCATCACCCCCG	120
OY	121	GCCGGCGTGGCGGAATAACCCGTGGTGGCATGAGTGGCTCATCCAGGAGGTGCGGA	180
Dd	121	GCCGGCGTGGCGGAATAACCCGTGGTGGCATGAGTGGCTCATCCAGGAGGTGCGGA	180
OY	181	GAGGGCTGGCTATACCGATCCGATTTCAAGCGGCGACGGCGCACTGTCGGCAC	240
Dd	181	GAGGGCTGGCTATACCGATCCGATTTCAAGCGGCGACGGCGCACTGTCGGCAC	240
OY	241	GCCGACAGGTGTGGGGCGAGCGCTATTTATGTCTCAAGTCAAGAACAAGATAGCGGG	300
Dd	241	GCCGACAGGTGTGGGGCGAGCGCTATTTATGTCTCAAGTCAAGAACAAGATAGCGGG	300
OY	301	GAATACGGCGCGCTCGAGACAGGGGAGATCTGTTCACGTTCTTGCAATTGGCCGCTCA	360
Dd	301	GAATACGGCGCGCTCGAGACAGGGGAGATCTGTTCACGTTCTTGCAATTGGCCGCTCA	360
OY	361	CGTGGTACCAGATGCGTTGTGGATTCGGACACCAAGTCAATTGCTACGAGACCGTC	420
Dd	361	CGTGGTACCAGATGCGTTGTGGATTCGGACACCAAGTCAATTGCTACGAGACCGTC	420
OY	421	CAGACCGCGAGCGGACATACCCCGTGGCGCCGATAGAGGAAGTCCCGGTGACATC	480
Dd	421	CAGACCGCGAGCGGACATACCCCGTGGCGCCGATAGAGGAAGTCCCGGTGACATC	480
OY	481	GCCGCCAGAGTTGGCGCTTACCACTGATGCGAACCAAGGGGGCGCGGTGTCTGATG	540
Dd	481	GCCGCCAGAGTTGGCGCTTACCACTGATGCGAACCAAGGGGGCGCGGTGTCTGATG	540
OY	541	GCGGGGGTCCGGGGTGCACCGGCGAGTGTGTATGGGGCGGACCGCGCGC	600
Dd	541	GCGGGGGTCCGGGGTGCACCGGCGAGTGTGTATGGGGCGGACCGCGCGC	600

QY 601 TACACGACGCGCATCGCCAAAGCGCATGGGCGGCGACCGCTTACGGTTCTAGACATCAAC 660  
|||||  
Db 601 TACACGACGCGCGCATCGCCAAAGCGCATGGGCGGCGACCGCTTACGGTTCTAGACATCAAC 660  
QY 661 ATGACAACTTGGGCACTGCGACGCGGAGTTCTGGGCGGAGTCCACACTCGCTACTCA 720  
|||||  
Db 661 ATGACAACTTGGGCACTGCGACGCGGAGTTCTGGGCGGAGTCCACACTCGCTACTCA 720  
QY 721 TCGGCGTACGAGCTCGAGGGTGGCGTCAAAAGTGGCGGAGTGGGATGGGGCGGTCTG 780  
|||||  
Db 721 TCGGCGTACGAGCTCGAGGGTGGCGTCAAAAGTGGCGGAGTGGGATGGGGCGGTCTG 780  
QY 781 GTGCGAGCGCCCAAGGCGCAATTAATAGTCTGCAATTCAGTTCTGCGGCAATGAAACA 840  
|||||  
Db 781 GTGCGAGCGCCCAAGGCGCAATTAATAGTCTGCAATTCAGTTCTGCGGCAATGAAACA 840  
QY 841 GGTGCGTACTGGTGGATATAGCATGACACAGGGGCGGTCTTCAAGGCTCAAGCCG 900  
|||||  
Db 841 GGTGCGTACTGGTGGATATAGCATGACACAGGGGCGGTCTTCAAGGCTCAAGCCG 900  
QY 901 ACCACCTACGACGCGGAGCTTGGCGTACGACGACGCTGTTTACTGCGTGGCGAAC 960  
|||||  
Db 901 ACCACCTACGACGCGGAGCTTGGCGTACGACGACGCTGTTTACTGCGTGGCGAAC 960  
QY 961 ATGCGCGCTCGGTGCGGAAAGCTGACCTACGCGCTGACCAAGCGGACGATCGCTAT 1020  
|||||  
Db 961 ATGCGCGCTCGGTGCGGAAAGCTGACCTACGCGCTGACCAAGCGGACGATCGCTAT 1020  
QY 1021 GTGCTGAGTGTGCGGCAATAGTGTGGGCGGCGGCGGTGGATCGGCGACTAGCC 1080  
|||||  
Db 1021 GTGCTGAGTGTGCGGCAATAGTGTGGGCGGCGGCGGTGGATCGGCGACTAGCC 1080  
QY 1081 AAAGGCTTTTGAAGCGCAAGAGGGGCGTACTGTCCGAAGGGGTGGCGGACTGGGG 1140  
|||||  
Db 1081 AAAGGCTTTTGAAGCGCAAGAGGGGCGTACTGTCCGAAGGGGTGGCGGACTGGGG 1140  
QY 1141 GTGCGCTTACGACGCGGAGCGGAGGCTGTGCTGCTGACTCTGCGCGCTGTACGCGAG 1200  
|||||  
Db 1141 GTGCGCTTACGACGCGGAGCGGAGGCTGTGCTGCTGACTCTGCGCGCTGTACGCGAG 1200  
QY 1201 CACACNTCGGAGTAGAAGAGCATGATGTCGNC 1235  
|||||  
Db 1201 CACACNTCGGAGTAGAAGAGCATGATGTCGNC 1235  
RESULT 4  
LOCUS A89753 1236 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 10 from Patent WO9832862.  
ACCESSION A89753  
VERSION A89753.1 GI:6738287  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1236)  
AUTHORS Flohe, L. and Singh, M.  
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIU4 MARINUM  
JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
FEATURES  
source location/Qualifiers  
1..1236  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 236 a 395 c 385 g 220 t  
ORIGIN  
Query Match 99.8%; Score 1232; DB 5; Length 1236;  
Best Local Similarity 99.8%; Pired. No. 3..3e-169;  
Matches 1232; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATCTGCAATTAATCAACTTCTTCACTGACACTGAAGCGTAC AGTATGAGAGGGGTATATC 60

|||||  
Db 1 ATCTGCAATTAATCAACTTCTTCACTGACACTGAAGCGTACAGTATCGAGAGGGGTATATC 60  
QY 61 ATGCGGCTGGGATATTCGACCGAGACAAAACAAATTCGCGGTGGCATCACCCG 120  
|||||  
Db 61 ATGCGGCTGGGATATTCGACCGAGACAAAACAAATTCGCGGTGGCATCACCCG 120  
QY 121 GCCGCGTGGCGGAATTAACCGGTGTGGCATAGGAGTCTATCCAGGAGTGGCCGA 180  
|||||  
Db 121 GCCGCGTGGCGGAATTAACCGGTGTGGCATAGGAGTCTATCCAGGAGTGGCCGA 180  
QY 181 GAGGCGTGGGCTATCACGACGCGGATTTCAAGGCGGAGCGCGCACTGTGGCAC 240  
|||||  
Db 181 GAGGCGTGGGCTATCACGACGCGGATTTCAAGGCGGAGCGCGCACTGTGGCAC 240  
QY 241 GCCGACGAGTGTGGGCGGACGCTGATTTATGCTCAAGTCAAGAACCGATAGCGCG 300  
|||||  
Db 241 GCCGACGAGTGTGGGCGGACGCTGATTTATGCTCAAGTCAAGAACCGATAGCGCG 300  
QY 301 GAATACGCGCGCTGCGACACGCGGATCTTGTACGCTTGTGATTTGGCGCGTCA 360  
|||||  
Db 301 GAATACGCGCGCTGCGACACGCGGATCTTGTACGCTTGTGATTTGGCGCGTCA 360  
QY 361 CGTGTTCACGCGATCGTTGTGATTCGCGACACGTCATTTGCTCAAGTCAAGAACCGATAGCGCG 420  
|||||  
Db 361 CGTGTTCACGCGATCGTTGTGATTCGCGACACGTCATTTGCTCAAGTCAAGAACCGATAGCGCG 420  
QY 421 CAGACGCGGAGCGGCGACATACCGCTGCTGGCGGATAGGAAAGTGGCGGTCGAC 480  
|||||  
Db 421 CAGACGCGGAGCGGCGACATACCGCTGCTGGCGGATAGGAAAGTGGCGGTCGAC 480  
QY 481 GCGCGCGAGTGTGGCGCTTACCACTGATGCGAACCAAGGGGCGCGGTGTGCTGATG 540  
|||||  
Db 481 GCGCGCGAGTGTGGCGCTTACCACTGATGCGAACCAAGGGGCGCGGTGTGCTGATG 540  
QY 541 GCGGCGGTGCGCGGCGTGCACCGCGCGACGTCGTGTGATGCGGCGCGCACCGCGCG 600  
|||||  
Db 541 GCGGCGGTGCGCGGCGTGCACCGCGCGACGTCGTGTGATGCGGCGCGCACCGCGCG 600  
QY 601 TACACGACGCGCGCATGCGCAAGCGGATGGGCGGAGCTTACGGTTCTAGACATCAAC 660  
|||||  
Db 601 TACACGACGCGCGCATGCGCAAGCGGATGGGCGGAGCTTACGGTTCTAGACATCAAC 660  
QY 661 ATGACAACTTGGGCACTGCGACGCGGAGTTCTGGGCGGAGTCCACACTCGCTACTCA 720  
|||||  
Db 661 ATGACAACTTGGGCACTGCGACGCGGAGTTCTGGGCGGAGTCCACACTCGCTACTCA 720  
QY 721 TCGGCGTACGAGCTCGAGGGTGGCGTCAAAAGTGGCGGAGTGGGATGGGGCGGTCTG 780  
|||||  
Db 721 TCGGCGTACGAGCTCGAGGGTGGCGTCAAAAGTGGCGGAGTGGGATGGGGCGGTCTG 780  
QY 781 GTGCGAGCGCCCAAGGCGCAATTAATAGTCTGCAATTCAGTTCTGCGGCAATGAAACA 840  
|||||  
Db 781 GTGCGAGCGCCCAAGGCGCAATTAATAGTCTGCAATTCAGTTCTGCGGCAATGAAACA 840  
QY 841 GGTGCGTACTGGTGGATATAGCATGACACAGGGGCGGTCTTCAAGGCTCAAGCCG 900  
|||||  
Db 841 GGTGCGTACTGGTGGATATAGCATGACACAGGGGCGGTCTTCAAGGCTCAAGCCG 900  
QY 901 ACCACTACGACGCGGAGCTTGGCGTACGACGACGCTGTTTACTGCGTGGCGAAC 960  
|||||  
Db 901 ACCACTACGACGCGGAGCTTGGCGTACGACGACGCTGTTTACTGCGTGGCGAAC 960  
QY 961 ATGCGCGCTCGGTGCGGAAAGCTGACCTACGCGCTGACCAAGCGGACGATCGCTAT 1020  
|||||  
Db 961 ATGCGCGCTCGGTGCGGAAAGCTGACCTACGCGCTGACCAAGCGGACGATCGCTAT 1020  
QY 1021 GTGCTGAGTGTGCGGCAATAGTGTGGGCGGCGGCGGTGGATCGGCGACTAGCC 1080  
|||||  
Db 1021 GTGCTGAGTGTGCGGCAATAGTGTGGGCGGCGGCGGTGGATCGGCGACTAGCC 1080  
QY 1081 AAAGGCTTTTGAAGCGCAAGAGGGGCGTACTGTCCGAAGGGGTGGCGGACTGGGG 1140  
|||||

```

Db 1081 AAGGCTTTTCAGCAGCAGAGGGGGCTTACTGTCCGAGGGGTGGCCACCGACCTGGGG 1140
QY 1141 GTGCGCTTACCGAGCCCGGAGCGTGTCTGAGCTCTGCGCGGTCTGTTACCGGAG 1200
Db 1141 GTGCGCTTACCGAGCCCGGAGCGTGTCTGAGCTCTGCGCGGTCTGTTACCGGAG 1200
QY 1201 CACACNTCGGGAGTAANGAAGCGATGATGTCNC 1235
Db 1201 CACACGTGGGAGTAAGGAAGCGATGATGTCGCG 1235

RESULT 5
AB7607 1237 bp DNA PAT 22-JAN-2000
LOCUS A87607 Sequence 4 from Patent WO9836089.
DEFINITION A87607
ACCESSION A87607.1 GI:6736247
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1237)
AUTHORS Flohe L. and Singh M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
source 1..1237
Location/Qualifiers
1..1237
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 236 a 394 c 386 g 221 t
ORIGIN

Query Match 99.68; Score 1230.4; DB 5; Length 1237;
Best Local Similarity 99.78; Pired. No. 5; 6e-169;
Matches 1231; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCTTGAGATTATCGAATTTCTTCACTGAGAGGTACGATGAGAGGGTAAATC 60
Db 1 ATCTTGAGATTATCGAATTTCTTCACTGAGAGGTACGATGAGAGGGTAAATC 60
QY 61 ATGCGCGTGGTATTCGCGACGAGCAAAAAACAAGATTCGGGTGCCATACCCCG 120
Db 61 ATGCGCGTGGTATTCGCGACGAGCAAAAAACAAGATTCGGGTGCCATACCCCG 120
QY 121 GCGGCGCTGCGGAGCACTAACCCGCTGCGGCAATGAGTGTCTATCCAGGAGGTCGCGA 180
Db 121 GCGGCGCTGCGGAGCACTAACCCGCTGCGGCAATGAGTGTCTATCCAGGAGGTCGCGA 180
QY 181 GAGGCTCGGCTATACAGCGAGCGGATTTAAGCGCGAGCGCGCAACTGTGCGGACC 240
Db 181 GAGGCTCGGCTATACAGCGAGCGGATTTAAGCGCGAGCGCGCAACTGTGCGGACC 240
QY 241 GCGGACGAGTGTGCGCGAGCGCTGATTTATGCTCAAGGTCAAAAGAACGATAGCGCG 300
Db 241 GCGGACGAGTGTGCGCGAGCGCTGATTTATGCTCAAGGTCAAAAGAACGATAGCGCG 300
QY 301 GAATAGCGCGCTGAGAGACGCGGAGATCTGTTCAGTTCCTTTCGATTTGGCGCGCTCA 360
Db 301 GAATAGCGCGCTGAGAGACGCGGAGATCTGTTCAGTTCCTTTCGATTTGGCGCGCTCA 360
QY 361 CGTCTTGAACGAGTGTGATTCGCGAGCAAGTCAATTCGCTACGAGAACGCTC 420
Db 361 CGTCTTGAACGAGTGTGATTCGCGAGCAAGTCAATTCGCTACGAGAACGCTC 420
QY 421 CAGACGCGAGCGGAGCACTACCGCTTGCCTGCGGAGTGAAGTGGCGGCTGCACTC 480
Db 421 CAGACGCGAGCGGAGCACTACCGCTTGCCTGCGGAGTGAAGTGGCGGCTGCACTC 480
QY 481 GCGGCGCGAGTGTGCGCTTACCACTGATGCAACCAAGGAGGCGCGGTGCTGATG 540
Db 481 GCGGCGCGAGTGTGCGCTTACCACTGATGCAACCAAGGAGGCGCGGTGCTGATG 540

```

```

QY 541 GCGGCGTCCCGCGGCTGCAACCGGCGGACGTGTGTATTCGGCGCGGACCGCGCG 600
Db 541 GCGGCGTCCCGCGGCTGCAACCGGCGGACGTGTGTATTCGGCGCGGACCGCGCG 600
QY 601 TACAGCGACCGCGCATTCGCAACGGCATGCGGCGACGCTTACGCTTACATCAAC 660
Db 601 TACAGCGACCGCGCATTCGCAACGGCATGCGGCGACGCTTACGCTTACATCAAC 660
QY 661 ATGACCAATTTGGCACTGACGCGGAGTTCGCGCGCGGATCCACATGCTACTCA 720
Db 661 ATGACCAATTTGGCACTGACGCGGAGTTCGCGCGCGGATCCACATGCTACTCA 720
QY 721 TCGGCTCAGAGCTCGAGGCTGCGCTCAACGTCGCGACCTGATTTGGGCGCGCG 780
Db 721 TCGGCTCAGAGCTCGAGGCTGCGCTCAACGTCGCGACCTGATTTGGGCGCGCG 780
QY 781 GTGCGAGCGCGCAAGCGCACCAATTAATGCTCGAATTCATTTGCGCGCATATGAACA 840
Db 781 GTGCGAGCGCGCAAGCGCACCAATTAATGCTCGAATTCATTTGCGCGCATATGAACA 840
QY 841 GTGCGGTAATGATGATATGCAATGACGACGAGGCGGCTGTTCAAGGCTCAGACCG 900
Db 841 GTGCGGTAATGATGATATGCAATGACGACGAGGCGGCTGTTCAAGGCTCAGACCG 900
QY 901 ACCACCTACGACACCGCGAGCTTGGCGGCGAGCAGACGCTGTTTACGCTGGGAGC 960
Db 901 ACCACCTACGACACCGCGAGCTTGGCGGCGAGCAGACGCTGTTTACGCTGGGAGC 960
QY 961 ATGCGGCTCTGCGGCGCAAGAGCTGACCTACGCGCTGACCAACGCGAGTCCGTAT 1020
Db 961 ATGCGGCTCTGCGGCGCAAGAGCTGACCTACGCGCTGACCAACGCGAGTCCGTAT 1020
QY 1021 GTGCTGAGCTTCCGACCATGAGTGGCGGCGGCTGCGGCTGCAATCCGACATAGCC 1080
Db 1021 GTGCTGAGCTTCCGACCATGAGTGGCGGCGGCTGCGGCTGCAATCCGACATAGCC 1080
QY 1081 AAGGCTTTTCAGCAGCAGAGGGGGTAACTGTCCGAAGGGGTGACCGACCTGGGG 1140
Db 1081 AAGGCTTTTCAGCAGCAGAGGGGGTAACTGTCCGAAGGGGTGACCGACCTGGGG 1140
QY 1141 GTGCGCTTACCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
Db 1141 GTGCGCTTACCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
QY 1201 CACACNTCGGAGTAANGAAGCGATGATGTCNC 1235
Db 1201 CACACGTGGGAGTAAGGAAGCGATGATGTCGCG 1235

RESULT 6
AB9747 1237 bp DNA PAT 22-JAN-2000
LOCUS A89747 Sequence 4 from Patent WO9832852.
DEFINITION A89747
ACCESSION A89747
VERSION A89747.1 GI:6738281
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1237)
AUTHORS Flohe L. and Singh M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832852-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
source 1..1237
Location/Qualifiers
1..1237
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 236 a 394 c 386 g 221 t
ORIGIN

```



Query Match 99.6%; Score 1230.4; DB 5; Length 1237;  
 Best Local Similarity 99.7%; Pred. No. 5,6e-169;  
 Matches 1231; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 ATCTTGAGATTATGCAACTTTCTTACACTGAGGCTACAGTATGAGAGGGTATC 60
    |||
DB 1 ATCTTGAGATTATGCAACTTTCTTACACTGAGGCTACAGTATGAGAGGGTATC 60

QY 61 ATGGGGCTGGTATTCGACGAGACCAAAACAAAGATTCCGGGTGGCATCACCCG 120
    |||
DB 61 ATGGGGCTGGTATTCGACGAGACCAAAACAAAGATTCCGGGTGGCATCACCCG 120

QY 121 GCGGGCTCGGGAACCTACCCGCTGCTGAGCATAGGTGCTCATCCAGGAGTGGCGA 180
    |||
DB 121 GCGGGCTCGGGAACCTACCCGCTGCTGAGCATAGGTGCTCATCCAGGAGTGGCGA 180

QY 181 GAGGGCTGGCTATACCGAGCGGGATTCAAGCGGGGAGCGGCGCACTGGTGGACC 240
    |||
DB 181 GAGGGCTGGCTATACCGAGCGGGATTCAAGCGGGGAGCGGCGCACTGGTGGACC 240

QY 241 GCGGACGAGTGTGGGCGGAGCGCTGATTTATGCTCAAGGTCAAAAGACCGATAGCGGC 300
    |||
DB 241 GCGGACGAGTGTGGGCGGAGCGCTGATTTATGCTCAAGGTCAAAAGACCGATAGCGGC 300

QY 301 GAATACGCGCGCTGCGACACAGGGGCAATCTTGTTCAGTTCCTTGATGGCT 360
    |||
DB 301 GAATACGCGCGCTGCGACACAGGGGCAATCTTGTTCAGTTCCTTGATGGCT 360

QY 361 CGGCTGACGCGATGCTGTTGGATTGCGGCGACGACGATTCCTTCGAACCGTGC 420
    |||
DB 361 CGGCTGACGCGATGCTGTTGGATTGCGGCGACGACGATTCCTTCGAACCGTGC 420

QY 421 CAGACCGCGACGCGGCGACTACCTGCTTGCCTCGATGAGCGAAGTGGCGGTGCACTC 480
    |||
DB 421 CAGACCGCGACGCGGCGACTACCTGCTTGCCTCGATGAGCGAAGTGGCGGTGCACTC 480

QY 481 GCGGCGCGAGTGGCGCTTACACCTGATGCGAAGCCAGGGGCGCGGTGTGCTGATG 540
    |||
DB 481 GCGGCGCGAGTGGCGCTTACACCTGATGCGAAGCCAGGGGCGCGGTGTGCTGATG 540

QY 541 GGGGGGTGGCGGCGCTGCAAGCGGGGCGAGTGGTGTGATGGGCGCGGCGCGCGGC 600
    |||
DB 541 GGGGGGTGGCGGCGCTGCAAGCGGGGCGAGTGGTGTGATGGGCGCGGCGCGCGGC 600

QY 601 TACAGCGACCGCGCATCGCCAAAGGCGATGGCGCGACCGTTACGTTCTAGACATCAAC 660
    |||
DB 601 TACAGCGACCGCGCATCGCCAAAGGCGATGGCGCGACCGTTACGTTCTAGACATCAAC 660

QY 661 ATGACGAAATTTGGGCACTGCACTGCGGCGGCGGATCCACACTCGCTACTCA 720
    |||
DB 661 ATGACGAAATTTGGGCACTGCACTGCGGCGGCGGATCCACACTCGCTACTCA 720

QY 721 TCGGCGCTACGAGTGCAGGGTGGCGTCAAAAGTSCGACCTGGTGATGGGCGGTCTG 780
    |||
DB 721 TCGGCGCTACGAGTGCAGGGTGGCGTCAAAAGTSCGACCTGGTGATGGGCGGTCTG 780

QY 781 GTGCCAGGCGCCAAAGGCGCAAAATAGTCTCGAATCACTGTGCGGCATATGAACCA 840
    |||
DB 781 GTGCCAGGCGCCAAAGGCGCAAAATAGTCTCGAATCACTGTGCGGCATATGAACCA 840

QY 841 GGTGGCGTACTGTGTGATATAGCCATGACAGAGGCGGCTTTTCAAGGCTCACAGCG 900
    |||
DB 841 GGTGGCGTACTGTGTGATATAGCCATGACAGAGGCGGCTTTTCAAGGCTCACAGCG 900

QY 901 ACCACCTACACACCGGAGCTTGCCTGACAGACAGCTGTTTACTCGTGGCGAAC 960
    |||
DB 901 ACCACCTACACACCGGAGCTTGCCTGACAGACAGCTGTTTACTCGTGGCGAAC 960

QY 961 ATGCGCGCTCGTGTGCGGAAGAGCTGACCTAGCGCTGACCAAGCGGAGATGCCGTAT 1020
    |||
DB 961 ATGCGCGCTCGTGTGCGGAAGAGCTGACCTAGCGCTGACCAAGCGGAGATGCCGTAT 1020

QY 1021 GTGCTGAGCTTGCGCGACATGGCTGGCGGGCGGTGCGGTCGAATCCGCGACTAGCC 1080

```

```

DB 1021 GTGCTGAGCTTGCGCGACATGGCTGGCGGGCGGTGCGGTCGAATCCGCGACTAGCC 1080
    |||
QY 1081 AAAGGTCTTTCGACGACGAGGCGGCTTACTGTCCGAAGGGTGGCCACGACCTGGGG 1140
    |||
DB 1081 AAAGGTCTTTCGACGACGAGGCGGCTTACTGTCCGAAGGGTGGCCACGACCTGGGG 1140

QY 1141 GTGCGTTCACGAGCGCGCGCGTGGCTGACCTGTCGCGGCTGTATACCGCGAG 1200
    |||
DB 1141 GTGCGTTCACGAGCGCGCGCGTGGCTGACCTGTCGCGGCTGTATACCGCGAG 1200

QY 1201 CACACNTCGGAGTAANGAAGCATGATGTCGMC 1235
    |||
DB 1201 CACACNTCGGAGTAANGAAGCATGATGTCGMC 1235

RESULT 7
MTV002 56414 bp DNA BCT 17-JUN-1998
LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.
DEFINITION AL008967 AL123456
ACCESSION AL008967.1 GI:3261491
VERSION
KEYWORDS
SOURCE Mycobacterium tuberculosis.
ORGANISM Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium.
REFERENCE 1 (bases 1 to 56414)
  Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
  Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E.,
  Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
  Connor,R., Davies,R., Devlin,K., Felwell,T., Gentles,S.,
  Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
  Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
  Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
  Squares,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and
  Barrell,B.G.
  Deciphering the biology of Mycobacterium tuberculosis from the
  complete genome sequence
  Nature 393 (6685), 537-544 (1998)
  98295987
  ERRATUM: [[published erratum appears in Nature 1998 Nov
  12;396(6707):190]]
  2 (bases 1 to 56414)
  Parkhill,J.
  Direct Submission
  Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
  tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
  Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
  Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
  75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
  On Jun 27, 1998 this sequence version replaced g1:2642326.
  NOTES:
  Details of M. tuberculosis sequencing at the Sanger Centre are
  available on the World Wide Web.
  (URL, http://www.sanger.ac.uk/Projects/M.tuberculosis/) CDS have
  been renumbered from the original cosmid submissions but the old
  gene designations are in brackets after the new gene numbers.
  Gene prediction was based on a Hidden Markov Model of TB genes
  implemented in Tlpase (Kiroh) supplemented with visual inspection
  of positional base preference in codons, especially where there is
  an increase in the observed/expected third position G + C.
  CAUTION: In some cases we may not have predicted the correct
  initiation codon. Where possible we choose an initiation codon
  (atg, gtg, or ttg) which is preceded by an upstream ribosome
  binding site sequence (optimally 5-11bp before the initiation
  codon). If this cannot be identified we choose the most upstream
  initiation codon.
  FEATURES
  source location/Qualifiers
  1..56414
  /organism="Mycobacterium tuberculosis"
  /strain="H37Rv"

```

```

misc_feature      <db_xref="taxon:1773"
                  <1..56414
                  /note="fragment designated v002. Does not represent a
                  physical clone"
source            /organism="Mycobacterium tuberculosis"
                  /strain="H37Rv"
                  /db_xref="taxon:1773"
                  /clone="Y154"
                  complement(3..527)
                  /gene="recA"
                  complement(3..527)
                  /gene="recX"
                  /note="RV2736c, (MTV002.01c), len: 174 aa. recX, similar
                  to eg. RECX_PSPAE_P37860 regulatory protein recX from
                  Pseudomonas aeruginosa (153 aa), fasta scores: opt: 161
                  z-score: 257.2 E(): 3.6e-07, 30.7% identity in 137 aa
                  overlap. Overlaps and extends CDS from overlapping cosmid
                  MTCY154.16c"
                  /codon_start=1
                  /transl_table=11
                  /product="recX"
                  /protein_id="CA15532.1"
                  /db_xref="GI:2624257"
                  /db_xref="SPTREMBL:O33280"
                  /translation="MTVSCPPSTSEREQARALCLRLTARSRTRELALQAKRG
                  PEDIGNRVLDRLAAVLVDVDFEOMQOSRRANAASKRALAELHAKGVDDVITV
                  VLGIGDAGERRAEKLVARLRREVLIDGTDENARSRRLVAMLARNGYQTLACEV
                  VIALAERERRRV"
                  complement(493..2865)
                  /gene="recA"
                  complement(493..2865)
                  /note="RV2737c, (MTV002.02c), len: 790 aa. recA,
                  identical to RECA_MYCTU_P26345 recA protein (790 aa).
                  Contains self-splicing protein element (intein) from 2294
                  to 969 (c), similar to intein II from TR_E332317
                  (EMBL:Y13030) DNA-directed DNA polymerase (EC 2.7.7.7) from
                  Thermococcus sp. (1829 aa), fasta scores: opt: 81 z-score:
                  235.2 E(): 6e-06, 24.6% identity in 183 aa overlap.
                  Contains PS00017 ATP/GTP-binding site motif A (P-loop),
                  PS00321 recA signature, and PS00881 protein splicing
                  signature. See Davis et al, (1992) Cell 117(12):201-210"
                  /codon_start=1
                  /transl_table=11
                  /product="recA"
                  /protein_id="CA15533.1"
                  /db_xref="GI:2624258"
                  /db_xref="SWISS-PROT:P26345"
                  /translation="MTQTPDREKALELAVAQIEKSGKSVRLDDEARQPISTVPTG
                  STADVALGIGLPGKRVLEIYGPESSGKTTVALAVANAQAGVAAFIDREHALDP
                  DYAKLGVDTDLVSPDTGOALEIDMLRSALDIVIDVSAALVPAVRALEKGM
                  GSHVGLARLMSQALRKMTGLNNNGTATFIRNLDRIQVMPSPETTGKALKE
                  YASVNDVRRVETLKGCTNAGVNRTRVYKKNKLAEGRIPT VTGTHREDVVDG
                  RPIHYVAAKAGTILHARPVYMPDQCPDVGICLAIAGAIYMAIPDKHVLLEYGMR
                  AELRKDRVAQPRRDFGDSNPPIADRAELILGLIDGRGRCWGKGRPIEFIVOR
                  ALIDVTRIAATLCAHPOGRISLAIARIEGRGVADLCCOAGITGKLANEKIPN
                  NTFEPDIAIVGNLFGLESDGWVREQ FALVAGYTTSEQLAHQHLILRGGV
                  GSTVNDYDPTOKRPSIVNGRIQSKRQVRE FISMVDNTATAESV-ANGPFGALIC
                  AIPKATOGRRGSOATYLAEMTIDVNLDERGVATYAEAAAMIGVASGDPKRGKIO
                  LGASRLRRDVALDALDDKFLHMLAELRYSYIAREVLPTRARATFDLEELHTL
                  VAEVVVHNCSPPEKQAEFDILYKGISREGSLIDMGVQGLIRKSGAMFTVEGRLG
                  OKENAVRNFLVNAVDVADEIKKIKIGICAVVVDPSNDGVLPAVPDF"
                  complement(790..2112)
                  /gene="recA"
                  /note="recA intein"
                  complement(790..807)
                  /gene="recA"
                  /note="PS00881 Protein splicing signature"
                  complement(2197..2223)
                  /gene="recA"
                  /note="PS00321 recA signature"
                  complement(2644..2667)
                  misc_feature

RBS              /gene="recA"
                  /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
                  complement(2873..2877)
                  /note="possible RBS upstream of RV2737c"
                  complement(3247..3453)
                  /gene="RV2738c"
                  complement(3247..3453)
                  /gene="RV2738c"
                  complement(3247..3453)
                  /gene="RV2738c"
                  complement(3247..3453)
                  /note="RV2738c, (MTV002.03c), len: 68 aa. Unknown,
                  N-terminus is highly similar to the N-terminus of the
                  upstream ORF MTV002.07c (78.4% identity in 37 aa overlap);
                  also similar to AL020958|SC4H8.5 Streptomyces coelicolor
                  cosmid 4H8 (64aa) opt: 185 z-score: 283.5 E(): 2.9e-08;
                  39.7% identity in 63 aa overlap"
                  /codon_start=1
                  /transl_table=11
                  /product="hypothetical protein RV2738c"
                  /protein_id="CA15534.1"
                  /db_xref="GI:2624260"
                  /db_xref="SPTREMBL:O33281"
                  /translation="MLAGVRLTFEHERVALHFGAAYGVSSVLLDHYLTGPDGSAQAQI
                  EDGVEPRDWRALCADFVPHDRW"
                  complement(3464..4630)
                  /gene="RV2739c"
                  complement(3464..4630)
                  /gene="RV2739c"
                  complement(3464..4630)
                  /note="RV2739c, (MTV002.04c), len: 388 aa. Probable
                  transferase, similar to eg. TR_Q31560 (EMBL:L28170)
                  Pseudomonas aeruginosa rhamnosyl transferase (426 aa),
                  fasta scores: opt: 178 z-score: 226.3 E(): 1.9e-05, 25.9%
                  identity in 425 aa overlap. Equivalent to Mycobacterium
                  leprae protein MLCB33.02c (399 aa); fasta scores
                  gpi1294723|MLCB33.2 Mycobacterium leprae cosmid B33 opt:
                  2112 z-score: 2364.5 E(): 0; 80.9% identity in 388 aa
                  overlap"
                  /codon_start=1
                  /transl_table=11
                  /product="hypothetical protein RV2739c"
                  /protein_id="CA15535.1"
                  /db_xref="GI:2624261"
                  /db_xref="SPTREMBL:O33282"
                  /translation="MRYAVAGPDPGHSFPALALCFRRAADPTLTFTGEMLEAR
                  AAGIDVDELGLAIDRLDAGARIHRAQOMVNVNPRALREPLVSDVDTAGG
                  MAEVLGIPWELNRPPLYPYLSKGLPIGSLAGCIRGRADATMRALTGSMRAG
                  LRQRAARVEICLPANRRCVTLIALPLALEPRPMFEAVVGVPLGEPDRTVLA
                  IPACTGVVYVASTALNCTAGLFEVALGSLTGTGTPYSGSRSLVSRSGADLVPPA
                  AVAGLGSQAEELTRADLVICGGHGVAKTLLAGVPMVYPGGGDMETANRVYRGS
                  AVLIRPLTADALVAANVELSPFRFRAARAASVAGADPVRCADLALAG"
                  complement(4674..5123)
                  /gene="RV2740"
                  complement(4674..5123)
                  /gene="RV2740"
                  complement(4674..5123)
                  /note="RV2740, (MTV002.05), len: 149 aa. unknown, similar
                  to M. leprae hypothetical protein TR_Q4850 (EMBL:Z294722)
                  MLCB33.03 (B2235_F3_140) (178 aa), fasta scores: opt: 438
                  z-score: 538.9 E(): 7.3e-23, 51.6% identity in 161 aa
                  overlap"
                  /codon_start=1
                  /transl_table=11
                  /product="hypothetical protein RV2740"
                  /protein_id="CA15536.1"
                  /db_xref="GI:2624262"
                  /db_xref="SPTREMBL:O33283"
                  /translation="MAELTSPPTPTPTTAIRAVAFNLNALQNEFDVDAAGDDC
                  VYENVGSRIGRRATLTLRRMGVGFVKIRICADAATLTERDALLITIGPLRV
                  OFWCVGVFEVDDGRITILRDYFDVYDMFKGLRGLVALVPSKATL"
                  complement(5355..6932)
                  /gene="PE_PGRS"
                  complement(5355..6932)
                  /gene="PE_PGRS"
                  complement(5355..6932)
                  /note="RV2741, (MTV002.06), len: 525 aa; Member of M.
                  tuberculosis PE_PGRS subfamily, similar to many eg.

```

Query Match	99.6%	Score 1230.4;	DB 1;	Length 56414;
Best Local Similarity	99.7%	Pred. No. 2.4e-169;		
Matches 1231;	Conservative	0;	Mismatches 4;	Indels 0;
				Gaps 0

QY	1	ATCTTCGAGATTAAACGAACTTCTTCAACAGTGAAGCGCTACAGTATCGAGAGGGGTAATC	60
Db	38201	ATCTTCGAGATTAAACGAACTTCTTCAATCTTCTTCACTGAGAGCTACAGTATCGAGAGGGGTAATC	38260
QY	61	ATGGCGCTGCGATTTCGACCGAGACGACAAAACAAACGAAATTCGGGTTGGCATCTACCCCG	120
Db	38261	ATGGCGCTGCGATTTCGACCGAGACGACAAAACAAACGAAATTCGGGTTGGCATCTACCCCG	38320
QY	121	GCCGCGCTGCGGGAACCTAACCCGCTGTGGCCATGAGTGTCTATCCAGGACAGTCCGGA	180
Db	38321	GCCGCGCTGCGGGAACCTAACCCGCTGTGGCCATGAGTGTCTATCCAGGACAGTGTCCGGA	38380
QY	181	GAGGCTGGGCTATACCGAGCGCGGATTTCAAGGGGGAGCGGGCACTGGTGGGGACC	240
Db	38381	GAGGCTGGGCTATACCGAGCGCGGATTTCAAGGGGGAGCGGGCACTGGTGGGGACC	38440
QY	241	GCCGACCAAGGTGTGGGCGGACGCTGATTTATTGCTCAAGTCAAAAGCAACGATACGGCG	300
Db	38441	GCCGACCAAGGTGTGGGCGGACGCTGATTTATTGCTCAAGTCAAAAGCAACGATACGGCG	38500
QY	301	GAATAGGGCGGCTCGAGACAGGGGAGATCTGTTCACGTTCTTGATTTGGCCGCTCA	360
Db	38501	GAATAGGGCGGCTCGAGACAGGGGAGATCTGTTCACGTTCTTGATTTGGCCGCTCA	38560
QY	361	CGTGTTCACCGATGCGTTGTGTGATTCGGACACCACGTCATATTGCTACGAGACCGTC	420
Db	38561	CGTGTTCACCGATGCGTTGTGTGATTCGGACACCACGTCATATTGCTACGAGACCGTC	38620
QY	421	CAGACCGCCGATGGGGGCACTTACCCCGTGTGGCCCGATAGGGAAGTCCGCGTGACATC	480
Db	38621	CAGACCGCCGATGGGGGCACTTACCCCGTGTGGCCCGATAGGGAAGTCCGCGTGACATC	38680
QY	481	GCCGCGCAAGTGTGGGCTTACCACTGATGCAACCCAAAGGGGGCGCGGTGTCTGATG	540
Db	38681	GCCGCGCAAGTGTGGGCTTACCACTGATGCAACCCAAAGGGGGCGCGGTGTCTGATG	38740
QY	541	GGCGGGGTGCGCGGGGTGCAACCGGCGGACGTCGTGTATGCGGGCGGGCACCGCGGCG	600
Db	38741	GGCGGGGTGCGCGGGGTGCAACCGGCGGACGTCGTGTATGCGGGCGGGCACCGCGGCG	38800
QY	601	TACAAACGACGCCGCGATCGCAACGAGCATGGCGCGACCGTTACGGTTCTAGACATCAAC	660
Db	38801	TACAAACGACGCCGCGATCGCAACGAGCATGGCGCGACCGTTACGGTTCTAGACATCAAC	38860
QY	661	ATCGACAAACTTCGGCACTCGACGCGCGAGTTCTGGGGCGGATCCACACTGGCTACTCA	720
Db	38861	ATCGACAAACTTCGGCACTCGACGCGCGAGTTCTGGGGCGGATCCACACTGGCTACTCA	38920
QY	721	TGCGGCTACGAGCTCGAGGGTGCCTGCAACGCTGCGGACCTGGTGAATTTGGGGCCGTCTG	780
Db	38921	TGCGGCTACGAGCTCGAGGGTGCCTGCAACGCTGCGGACCTGGTGAATTTGGGGCCGTCTG	38980
QY	781	GTCGCCAGCGGCCAAGGCAACCCAAATTAAGTCTGGAATTACTGTGCGGCTATGAAACCA	840
Db	38981	GTCGCCAGCGGCCAAGGCAACCCAAATTAAGTCTGGAATTACTGTGCGGCTATGAAACCA	39040
QY	841	GGTGCGTACTGGTGAATATAGCATCGACAGGGGGGCTGTTTCAMAAGGCTACAGACCG	900
Db	39041	GGTGCGTACTGGTGAATATAGCATCGACAGGGGGGCTGTTTCAMAAGGCTACAGACCG	39100
QY	901	ACCACCTACGACCAACCCGACGTTCCGCGTCGACGACAGACGCTGTTTACTGTGCGGGAC	960
Db	39101	ACCACCTACGACCAACCCGACGTTCCGCGTCGACGACAGACGCTGTTTACTGTGCGGGAC	39160
QY	961	ATGCGCGGCTGGTGGCGGAAGAGCTGCACCTACGCGCTACCAACGACGAGATGCCGTAT	1020
Db	39161	ATGCGCGGCTGGTGGCGGAAGAGCTGCACCTACGCGCTACCAACGACGAGATGCCGTAT	39220
QY	1021	GTGCTCGACCTTGCAGACCATGCGTGGCGGGCGGCTGCCGGTCAATCCGGCACTAGCC	1080

Db	39221	GTGTCAGAGTTCCGACCATGGTGGCGGGGGGTGCCGGTGGATCCGGAATAGCC	39280
Qy	1081	AAAGCTCTTCGACGACGAAGGGCGTTACTGTCCGAACGGTGGCCACCGACTGGGG	1140
Db	39281	AAAGGTTTTCGACGACGAAGGGCGTTACTGTCCGAACGGTGGCCACCGACTGGGG	39340
Qy	1141	GTGCGGTTCCACGAGCCGCCAGCGTGTGGCGCTGTGACTGTGGCGCGCTGTACGCCAG	1200
Db	39341	GTGCGGTTCCACGAGCCGCCAGCGTGTGGCGCTGTGACTGTGCGCGCTGTACGCCAG	39400
Qy	1201	CACACNTGGGAGTAANGACAAGCATGATGTCGNC	1235
Db	39401	CACACGTCGGAGTAANGACAAGCATGATGTGCGC	39435

RESULT	8	1228 bp	DNA	PAT	22-JAN-2000
LOCUS	A87608	1228 bp	DNA		
DEFINITION	Sequence 5 from Patent WO936089.				
ACCESSION	A87608				
VERSION	A87608.1	GI:6736248			
KEYWORDS					
SOURCE	unidentified.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 1228)				
AUTHORS	Flohe, L. and Singh, M.				
TITLE	TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE				
JOURNAL	Patent: WO 9836089-A 20-ANG-1998; FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)				
FEATURES	Location/Qualifiers				
source	1..1228				
BASE COUNT	236 a 391 c 382 g 219 t				
ORIGIN	/organism="unidentified" /db_xref="taxon:32644"				
Query Match	99.1%; Score 1224.4; DB 5; Length 1228;				
Best Local Similarity	99.8%; Pred. No. 4.1e-166;				
Matches 1225: Conservative	0; Mismatches 3; Indels 0; Gaps 0;				
QY	1	ATCTTCGACATTAATCGAACTTTCTTCACACTGAGAGCGTATGTCGAGAGGGTAATC	60		
DB	1	ATCTTCGACATTAATCGAACTTTCTTCATCTAGAACCGTACAGTATCGAGAGGGTAATC	60		
QY	61	ATGCGGCTGGTATTCCGATCGACACCAAAAAACAACGAATTCGGGTGGCCATCACCCCG	120		
DB	61	ATGCGGCTGGTATTCCGACCGACGACCAAAAAACAACGAATTCGGGTGGCCATCACCCCG	120		
QY	121	GCGGCGTGGCGGAAGTAACCCGTCGTGGCCATGAGTCTCATCGAGGAGTCCGGA	180		
DB	121	GCGGCGTGGCGGAAGTAACCCGTCGTGGCCATGAGTCTCATCGAGGAGTCCGGA	180		
QY	181	GAGGGCTCGGCTATCAACGACGCGGATTTCAAGCGGCGAGGCGCCAACTGTGCGCAC	240		
DB	181	GAGGGCTCGGCTATCAACGACGCGGATTTCAAGCGGCGAGGCGCCAACTGTGCGCAC	240		
QY	241	GCGGACCAAGTGTGGGCGACGCTGATTATTGTCTCAAGTCAAAAGAACCTATTAGCGGG	300		
DB	241	GCGGACCAAGTGTGGGCGACGCTGATTATTGTCTCAAGTCAAAAGAACCTATTAGCGGG	300		
QY	301	GAAATACGCGCGCTGCGACACGCGGACAGATCTTGTTACGTTCTTGCAATTTGGCCGCTCA	360		
DB	301	GAAATACGCGCGCTGCGACACGCGGACAGATCTTGTTACGTTCTTGCAATTTGGCCGCTCA	360		
QY	361	CGTGGTGCACGATGCGTGTGTGATTCGCGGACACACGTCATTAATGCTTACGAGACCGTC	420		
DB	361	CGTGGTGCACGATGCGTGTGTGATTCGCGGACACACGTCATTAATGCTTACGAGACCGTC	420		
QY	421	CAGACCGCGGACGCGCACTACCCCTGCTTGGCCCCGATAGGAAATGCGCGGTGACTC	480		
DB	421	CAGACCGCGGACGCGCACTACCCCTGCTTGGCCCCGATAGGAAATGCGCGGTGACTC	480		





JOURNAL Patent: WO 9832862-A 30-JUL-1998;

FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)

## FEATURES

source

Location/Qualifiers

1. 1235

/organism="unidentified"

/db\_xref="taxon:32644"

## BASE COUNT

235 a 395 c 384 g 220 t 1 others

## ORIGIN

Query Match 98.7%; Score 1219; DB 5; Length 1235;

Best Local Similarity 99.6%; Pred. No. 2,5e-167; Matches 1230; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 ATCTGCAATTAATGCACTTTCTTACACAGTGAAGCTGATGATGAGGGGTAATC 60  
 DB 1 ATCTGCAATTAATGCACTTTCTTACACAGTGAAGCTGATGATGAGGGGTAATC 60  
 QY 61 ATGCGGCTGCTATTCGACCGAGACAAACAAACAAATTCGGGGTGC 3CATCACCCG 120  
 DB 61 ATGCGGCTGCTATTCGACCGAGACAAACAAACAAATTCGGGGTGC 3CATCACCCG 120  
 QY 121 GCGGCGTGGGCACTAAACCGCTGCTGCGCATGAGTGTCTATCCAGCAGTGC 180  
 DB 121 GCGGCGTGGGCACTAAACCGCTGCTGCGCATGAGTGTCTATCCAGCAGTGC 180  
 QY 121 GCGGCGTGGGCACTAAACCGCTGCTGCGCATGAGTGTCTATCCAGCAGTGC 180  
 DB 121 GCGGCGTGGGCACTAAACCGCTGCTGCGCATGAGTGTCTATCCAGCAGTGC 180  
 QY 181 GAGGCGTGGGCACTAAACCGCTGCTGCGCATGAGTGTCTATCCAGCAGTGC 240  
 DB 181 GAGGCGTGGGCACTAAACCGCTGCTGCGCATGAGTGTCTATCCAGCAGTGC 240  
 QY 241 GCGGACAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
 DB 241 GCGGACAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
 QY 301 GAATAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
 DB 301 GAATAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
 QY 361 CGAGCTTGAACGATGCTGTTGATGATCGGACGACGATTCCTTACGAGACG 420  
 DB 361 CGAGCTTGAACGATGCTGTTGATGATCGGACGACGATTCCTTACGAGACG 420  
 QY 421 CAGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480  
 DB 421 CAGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480  
 QY 481 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540  
 DB 481 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540  
 QY 541 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
 DB 541 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
 QY 601 TACAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660  
 DB 601 TACAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660  
 QY 661 ATGCAAAATTTGGGCACTGCAAGCGGCGGCGGCGGCGGCGGCGGCGG 720  
 DB 661 ATGCAAAATTTGGGCACTGCAAGCGGCGGCGGCGGCGGCGGCGGCGG 720  
 QY 721 TCGGCGTGAAGCTGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780  
 DB 721 TCGGCGTGAAGCTGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780  
 QY 781 GTCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840  
 DB 781 GTCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840  
 QY 841 GTCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
 DB 841 GTCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
 QY 840 GTCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 899  
 DB 840 GTCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 899

QY 901 ACCACTAGACACCCGACGTTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960  
 DB 901 ACCACTAGACACCCGACGTTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960  
 QY 961 ATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
 DB 961 ATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
 QY 1021 GTCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080  
 DB 1021 GTCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080  
 QY 1081 GTCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140  
 DB 1081 GTCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140  
 QY 1141 GTCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
 DB 1141 GTCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
 QY 1201 CACACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260  
 DB 1201 CACACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260

## RESULT 12

A87611

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 1230; Conservative

0; Mismatches

4; Indels

1; Gaps

1;

QY 1 ATCTGCAATTAATGCACTTTCTTACACAGTGAAGCTGATGATGAGGGGTAATC 60  
 DB 1 ATCTGCAATTAATGCACTTTCTTACACAGTGAAGCTGATGATGAGGGGTAATC 60  
 QY 61 ATGCGGCTGCTATTCGACCGAGACAAACAAACAAATTCGGGGTGC 3CATCACCCG 120  
 DB 61 ATGCGGCTGCTATTCGACCGAGACAAACAAACAAATTCGGGGTGC 3CATCACCCG 120  
 QY 121 GCGGCGTGGGCACTAAACCGCTGCTGCGCATGAGTGTCTATCCAGCAGTGC 180  
 DB 121 GCGGCGTGGGCACTAAACCGCTGCTGCGCATGAGTGTCTATCCAGCAGTGC 180  
 QY 181 GAGGCGTGGGCACTAAACCGCTGCTGCGCATGAGTGTCTATCCAGCAGTGC 240  
 DB 181 GAGGCGTGGGCACTAAACCGCTGCTGCGCATGAGTGTCTATCCAGCAGTGC 240  
 QY 241 GCGGACAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
 DB 241 GCGGACAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300







Db 780 GTCCAGGCGCCGAAGGACCCAAATGATCTGAAATTCATCTTGTCCGCATATGAAACA 839  
 QY 841 GGTGCGGACTGTGATATATGACATGACACGAGGCGGCTGTTTGAAGGCTCAGACCG 900  
 Db 840 GGTGCGGACTGTGATATATGACATGACACGAGGCGGCTGTTTGAAGGCTCAGACCG 899  
 QY 901 ACACACTACGACACCCGACGCTTGGCGTGCACGACACGCTGTTTACTGCGGCGAAC 960  
 Db 900 ACACACTACGACACCCGACGCTTGGCGTGCACGACACGCTGTTTACTGCGGCGAAC 959  
 QY 961 ATGCGCGCGCTGCGGCGGAAACGCTGACCTACCGCGCTGACACACGCGACATCCGAT 1020  
 Db 960 ATGCGCGCGCTGCGGCGGAAACGCTGACCTACCGCGCTGACACACGCGACATCCGAT 1019  
 QY 1021 GTGCTGAGCTTGGCGACATGCTGCGGCGGCGGCGGCTGCGGCTGAAATCCGCGACTAGCC 1080  
 Db 1020 GTGCTGAGCTTGGCGACATGCTGCGGCGGCGGCGGCTGCGGCTGAAATCCGCGACTAGCC 1079  
 QY 1081 AAAGGCTTTTCGACGACGACGAGGGGCGCTTACTGTCCGAAACGCGTGGCCACCGACCTGAGG 1140  
 Db 1080 AAAGGCTTTTCGACGACGACGAGGGGCGCTTACTGTCCGAAACGCGTGGCCACCGACCTGAGG 1139  
 QY 1141 GTGCGCTTACCGACGACGCGCGCGCGGCTGCTGCGCTGACTCTTGGCGCGCTGTTAGCGGAG 1200  
 Db 1140 GTGCGCTTACCGACGACGCGCGCGCGGCTGCTGCGCTGACTCTTGGCGCGCTGTTAGCGGAG 1199  
 QY 1201 CACACNTCGGAGTAAAGGAAAGCATGATGTCGNC 1235  
 Db 1200 CACACNTCGGAGTAAAGGAAAGCATGATGTCGNC 1234

RESULT 14  
 LOCUS A87605 1245 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 2 from Patent W09836089.  
 ACCESSION A87605  
 VERSION A87605.1 GI:6736245  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1245)  
 AUTHORS Flohe, L. and Singh, M.  
 TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source Location/Qualifiers  
 1..1245  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 238 a 398 c 387 g 222 t  
 ORIGIN

Query Match 98.5%; Score 1216; DB 5; Length 1245;  
 Best Local Similarity 99.3%; Pred. No. 6,7e-167;  
 Matches 1232; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY 1 ATCTTGAGATTATGCACTTTCTTCACTGAAGCTACAGTATGAGAGGGGTATC 60  
 Db 1 ATCTTGAGATTATGCACTTTCTTCACTGAAGCTACAGTATGAGAGGGGTATC 60  
 QY 61 ATGCGCGCTGATTTGCGACGAGAACCAAAACG-----AATTCGAGGTGCGCATC 114  
 Db 61 ATGCGCGCTGATTTGCGACGAGAACCAAAACGAAATTCGAATTCGAGGTGCGCATC 120  
 QY 115 ACCCGGCGCGCGCTGCGGAACTAACCGCTGCGCATGAGTGCATCCAGCAGGT 174  
 Db 121 ACCCGGCGCGCGCTGCGGAACTAACCGCTGCGCATGAGTGCATCCAGCAGGT 180  
 QY 175 GCGGAGAGGGCTGCGCTATCAGCAGCGGATTTCAAGCGGCGGCGCAACTGTC 234  
 Db 181 GCGGAGAGGGCTGCGCTATCAGCAGCGGATTTCAAGCGGCGGCGCAACTGTC 240

QY 235 GGCACCGCCGACAGGCTGCGGCGGCGGCGGCTATTATGCTCAAGGTCAAAGACCGATA 294  
 Db 241 GGCACCGCCGACAGGCTGCGGCGGCGGCGGCTATTATGCTCAAGGTCAAAGACCGATA 300  
 QY 295 GCGGCGGAATACGGCGCGCTGCGACACGGGCGAGATTTGTTACGCTTTCGATTTGGCC 354  
 Db 301 GCGGCGGAATACGGCGCGCTGCGACACGGGCGAGATTTGTTACGCTTTCGATTTGGCC 360  
 QY 355 GCGTACGCTCTTGCACCGATGCTGTTGATTCGGGACACGATCAATTGCTTACGAG 414  
 Db 361 GCGTACGCTCTTGCACCGATGCTGTTGATTCGGGACACGATCAATTGCTTACGAG 420  
 QY 415 ACCGTCAGACCGCGCGGCGGCGGCTACCGCTGCGGCGGCGGCGGCTGAGGCGGCGGT 474  
 Db 421 ACCGTCAGACCGCGCGGCGGCGGCTACCGCTGCGGCGGCGGCGGCTGAGGCGGCGGT 480  
 QY 475 CGACTGCGCCCGCAGGTTGGCGCTTACCACTGATGCGAACCAGGAGGCGCGGCTGTG 534  
 Db 481 CGACTGCGCCCGCAGGTTGGCGCTTACCACTGATGCGAACCAGGAGGCGCGGCTGTG 540  
 QY 535 CTGATGGCGGGGTGCGCGCGCTGCGAACCGGCGGACGCTGCTGATCGGCGCGGACCC 594  
 Db 541 CTGATGGCGGGGTGCGCGCGCTGCGAACCGGCGGACGCTGCTGATCGGCGCGGACCC 600  
 QY 595 GCGGCGTACAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 654  
 Db 601 GCGGCGTACAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660  
 QY 655 ATCAACATCGACAACTTGGCACTCGACGCGGAGTTCGCGGCGGCGGCGGCGGCGGCGG 714  
 Db 661 ATCAACATCGACAACTTGGCACTCGACGCGGAGTTCGCGGCGGCGGCGGCGGCGGCGG 720  
 QY 715 TACTCATGCGCTTACGAGCTCGAGGTCGCGTCAACGTCGCGACCTGCTGATTTGGGCGC 774  
 Db 721 TACTCATGCGCTTACGAGCTCGAGGTCGCGTCAACGTCGCGACCTGCTGATTTGGGCGC 780  
 QY 775 GTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 834  
 Db 781 GTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840  
 QY 835 AAACAGGTGCGGCTGATGATATGACATGACGAGGCGGCGGCTTTCGAAGGCTCA 894  
 Db 841 AAACAGGTGCGGCTGATGATATGACATGACGAGGCGGCGGCTTTCGAAGGCTCA 900  
 QY 895 CGACCGACACTTACGACACGCGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 954  
 Db 901 CGACCGACACTTACGACACGCGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960  
 QY 955 GCGAATGCGCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1014  
 Db 961 GCGAATGCGCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020  
 QY 1015 CCGTATGCTGCTGAGCTTGGCGACATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1074  
 Db 1021 CCGTATGCTGCTGAGCTTGGCGACATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080  
 QY 1075 CTGCGCAAGGCTTTTCGACGACGAGGCGGCTTACTGTCCGAACGCGGTGCGCAC 1134  
 Db 1081 CTGCGCAAGGCTTTTCGACGACGAGGCGGCTTACTGTCCGAACGCGGTGCGCAC 1140  
 QY 1135 CTGCGGCGGCGGCTTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1194  
 Db 1141 CTGCGGCGGCGGCTTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
 QY 1195 GCGGAGCACACNTCGGAGATTAAGGAAAGCATGATGTCGNC 1235  
 Db 1201 GCGGAGCACACNTCGGAGATTAAGGAAAGCATGATGTCGNC 1241

RESULT 15  
 A89745 1245 bp DNA PAT 22-JAN-2000  
 LOCUS



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:05:34 ; Search time 458.59 seconds  
(without alignments)  
673.777 Million cell updates/sec

Title: US-09-362-485-3

Perfect score: 1235  
Sequence: 1 ATCTTGCAATTATTCGAC.....ANGAGCGATGATGTCGNC 1235

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311985 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: N\_Geneseq.36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1216	98.5	1245	1 V49626	Mycobacterium tube
2	1216	98.5	1260	1 V49610	Mycobacterium sp.
3	1216	98.5	1260	1 V49625	Mycobacterium tube
4	467.6	37.9	682	1 V49511	Mycobacterium mari
5	277.2	22.4	1125	1 N91423	Sequence of heat-r
6	212.4	17.2	28171	1 V52155	Streptococcus pneu
7	178.4	14.4	9280	1 V74442	Streptococcus aur
8	94.8	7.7	1074	1 T67971	H. pylori membrane
9	69.2	5.6	544	1 V43039	Streptococcus pneu
10	57.6	4.7	31096	1 V74370	Staphylococcus aur
11	50.8	4.1	390	1 Q21833	Randomising oligon
12	50.8	4.1	390	1 Q36859	PCR primer for 5'
13	50.4	4.1	1413	1 X34205	Mycobacterium spec
14	50.4	4.1	1722	1 X34206	Mycobacterium spec
15	50.4	4.1	13872	1 T68715	Streptomyces venez
16	49	4.0	1269	1 Q03065	Sequence encoding
17	47.6	3.9	3346	1 T93610	Mycobacterium tube
18	47.6	3.9	28598	1 T06769	Sorangium cellulos
19	47.6	3.9	28598	1 T89956	Sorangium cellulos
20	47.6	3.9	49377	1 V05287	The streptophen biosy
21	46.4	3.8	985	1 V44439	Mycobacterium tube
22	46.4	3.8	985	1 V64548	M. tuberculosis im
23	46	3.7	1620	1 Q22482	groEL-1 gene codin
24	46	3.7	2582	1 T73117	Actinoplanes sp. a
25	46	3.7	2668	1 Q22485	groEL-1 gene. Reco
26	46	3.7	17955	1 V56642	Actinoplanes sp. a
27	45.4	3.7	833	1 Q64203	snab gene encoding
28	45.4	3.7	5392	1 Q64201	Sequence compislin
29	42.4	3.4	1833	1 Q64206	snb gene encoding
30	42.4	3.4	2185	1 V84066	Clone P5-4 encodin
31	42.4	3.4	2186	1 Q99365	S. lividans protea
32	42.4	3.4	24379	1 T93095	Streptomyces fireo
33	42.4	3.4	24379	1 V25925	Streptomyces roseo
34	42.2	3.4	329	1 V44425	Mycobacterium tube

35	42.2	3.4	329	1 V64534	M. tuberculosis im
36	42	3.4	3946	1 T93610	Mycobacterium tube
37	42	3.4	4018	1 Q63879	Polyhydroxyalkanoa
38	41.8	3.4	882	1 V44403	Mycobacterium tube
39	41.8	3.4	882	1 V64512	M. tuberculosis im
40	41.6	3.4	12036	1 Q04668	FHA structural gen
41	41.2	3.3	2414	1 Q05926	Sequence encoding
42	41	3.3	29879	1 Q46806	eryA region of S.
43	40.8	3.3	2151	1 Q76252	HSV-2 protease, IC
44	40.8	3.3	2151	1 Q76261	HSV-2 protease/ICP
45	40.8	3.3	2472	1 Q84671	HSV-2 UL6 gene. N

## ALIGNMENTS

RESULT	1	
VS49626		
ID	V49626 standard; DNA; 1245 BP.	
AC	V49626;	
DT	20-NOV-1998 (first entry)	
DE	Mycobacterium tuberculosis	
KM	SS: Alanine dehydrogenase: tuberculosis; TB; Mycobacteria; L-alanine.	
OS	Mycobacterium tuberculosis.	
PN	W09836089.f2.	
PD	20-AUG-1998.	
PF	29-JAN-1998; E00483.	
PR	29-JAN-1997; EP-1013338.	
PA	(FLOH/) FLOHE L.	
PI	Flohe L. Hutter B. Kolk A. Singh M.	
DR	WPI: 98-457123/39.	
PT	Diagnosis of, e.g., tuberculosis from alanine dehydrogenase activity	
PT	- useful for, e.g., for diagnosis, differentiation of strains,	
PT	monitoring vaccination and identification of mycobacterial	
PT	inhibitors	
PS	Disclosure: Fig 3.19; 55pp; German.	
CC	The Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in	
CC	the production of kits for diagnosing tuberculosis (TB) and other	
CC	mycobacterial infections in humans or animals. Kits are used for direct	
CC	diagnosis of TB on clinical samples (e.g., body fluids) and can	
CC	differentiate between pathogenic and non-virulent strains, e.g., for	
CC	identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may	
CC	also be used to identify substances that inhibit mycobacteria, for	
CC	combating epidemics and for vaccination follow-up. Oligonucleotides	
CC	derived from AlaDH are used similarly in diagnostic hybridisation tests,	
CC	also for culture confirmation of isolated strains and for chromosome	
CC	fingerprinting to detect/differentiate between mycobacteria, and for	
CC	L-alanine-specific biotransformation reactions. AlaDH is an early	
CC	antigen, present extracellularly after only a few days of growth, making	
CC	it an ideal drug target.	
SQ	Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;	
Query Match	98.5%; Score 1216; DB 1; Length 1245;	
Best Local Similarity	99.3%; Pred. No. 1.2e+24;	
Matches 1332; Conservative	0; Mismatches 3; Indels 6; Gaps 1;	
QY	1 ATCTTGCAATTATTCGACGTAAGCGTACGATGCAAGGGTAATC 60	
DB	1 ATCTTGCAATTATTCGACGTAAGCGTACGATGCAAGGGTAATC 60	
QY	61 ATGCGCGTGTATTCGACGACGACCAAAACACG-----AATCCGGGTGCGCATC 114	
DB	61 ATGCGCGTGTATTCGACGACGACCAAAACACGAAATTCGATTCGCGGTGCGCATC 120	
QY	115 ACCCGCGCGCGTGTGCGGACGTAACCGGTGCGGATGAGTCTTATCCAGGCGAGT 174	
DB	121 ACCCGCGCGCGTGTGCGGACGTAACCGGTGCGGATGAGTCTTATCCAGGCGAGT 180	
QY	175 GCGGAGAGGCGTGTGCGGATGACGACGCGGATTCAGGGGCGGCGGCGCACTGGTC 234	
DB	181 GCGGAGAGGCGTGTGCGGATGACGACGCGGATTCAGGGGCGGCGGCGCACTGGTC 240	

```

OY 235 GGCACCGCCGACAGGTGTGGGCGGACGCTGATTTATGCTCAAGSTCAAGAACCGATA 294
DB 241 GGCACCGCCGACAGGTGTGGGCGGACGCTGATTTATGCTCAAGSTCAAGAACCGATA 300
OY 295 GCGGCGGAATACGGCCGCTGCGACACGGGAGATCTTTACAGTCTTTCGATTTGGCC 354
DB 301 GCGGCGGAATACGGCCGCTGCGACACGGGAGATCTTTACAGTCTTTCGATTTGGCC 360
OY 355 GGGTACGCTGTGACCGCATGGTGTGGATTCGGGACACGATTCGCTACGAG 414
DB 361 GGGTACGCTGTGACCGCATGGTGTGGATTCGGGACACGATTCGCTACGAG 420
OY 415 ACCGTCACAGCCGCGACGCGCACTACCTCTCTCTGCCCCGATGACGAGTCGCGGT 474
DB 421 ACCGTCACAGCCGCGACGCGCACTACCTCTCTCTGCCCCGATGACGAGTCGCGGT 480
OY 475 CGACTGCGCCCGCAGGTTGGCGTTACACCTGATCGAAACCGAGGGGCGCGGTGTG 534
DB 481 CGACTGCGCCCGCAGGTTGGCGTTACACCTGATCGAAACCGAGGGGCGCGGTGTG 540
OY 535 CTGATGGGCGGCGGCGCGGCGTGAACCGGCGGACGCTGTGTGATCGGCGCGGACCC 594
DB 541 CTGATGGGCGGCGGCGGCGGCGTGAACCGGCGGACGCTGTGTGATCGGCGCGGACCC 600
OY 595 GCGGCGTACAGCGACCGCGCATCGCCAAAGCGCATGGCGCGACCGTACGGTTAGAC 654
DB 601 GCGGCGTACAGCGACCGCGCATCGCCAAAGCGCATGGCGCGACCGTACGGTTAGAC 660
OY 655 ATCAACATGACACAAATTTGGGAACTGACGCGGAGTCTTGCGGCGGAGTCCACATCGC 714
DB 661 ATCAACATGACACAAATTTGGGAACTGACGCGGAGTCTTGCGGCGGAGTCCACATCGC 720
OY 715 TACTCATCGGCTACGAGGTGCGAGGCTGCGCAACGTCGACCTGTGTGATTTGGGCGC 774
DB 721 TACTCATCGGCTACGAGGTGCGAGGCTGCGCAACGTCGACCTGTGTGATTTGGGCGC 780
OY 775 GTCTGTGTCGCGACGGCGCCAAAGGACCCAAATTAATGTCGAATTCCTGTGCGCGCATAG 834
DB 781 GTCTGTGTCGCGACGGCGCCAAAGGACCCAAATTAATGTCGAATTCCTGTGCGCGCATAG 840
OY 835 AAACGAGTGGGCTACTGTGATATAGCATGACGAGGCGGCTGTTCGAAGGCTCA 894
DB 841 AAACGAGTGGGCTACTGTGATATAGCATGACGAGGCGGCTGTTCGAAGGCTCA 900
OY 895 CGACGACCACTACGACCAACCGGACGTTGCGCGTGACAGACAGCTGTATTTACTGCGTG 954
DB 901 CGACGACCACTACGACCAACCGGACGTTGCGCGTGACAGACAGCTGTATTTACTGCGTG 960
OY 955 GCGAATATCCCGGCTGCGTGGCGGAAGAGTGTGACTACGCGCTGACCAACGCGACGATG 1014
DB 961 GCGAATATCCCGGCTGCGTGGCGGAAGAGTGTGACTACGCGCTGACCAACGCGACGATG 1020
OY 1015 CCGTATGCTGTGAGGTTGCGGACATGAGTGGCGGCGGCTGCGGCTGCAATCGGCA 1074
DB 1021 CCGTATGCTGTGAGGTTGCGGACATGAGTGGCGGCGGCTGCGGCTGCAATCGGCA 1080
OY 1075 CTAGCCAAAGGCTTTTGAAGGACGAGAGGGGCTTACTGTTCGAAGCGGTGTGCGCAACG 1134
DB 1081 CTAGCCAAAGGCTTTTGAAGGACGAGAGGGGCTTACTGTTCGAAGCGGTGTGCGCAACG 1140
OY 1135 CTGGGGGTGCGCTTACCGACGAGCGCGGCGGCTGTGAGTCTGCGCGCTGTGTAC 1194
DB 1141 CTGGGGGTGCGCTTACCGACGAGCGCGGCGGCTGTGAGTCTGCGCGCTGTGTAC 1200
OY 1195 GCGGAGCACACNTCGGAGTAANGAAGGATGATGTCNC 1235
DB 1201 GCGGAGCACACNTCGGAGTAANGAAGGATGATGTCNC 1241

```

RESULT 2  
 ID V49510 standard; DNA: 1260 BP.  
 AC V49510;

DT 20-OCT-1998 (first entry)  
 DE Mycobacterium sp. Aladh DNA.  
 KM Alanine dehydrogenase; Aladh; ADH; diagnosis; tuberculosis; pathogen;  
 KM swimmers disease; vaccine; epidemic; infection; identification; ss.  
 OS Mycobacterium sp.  
 PN M09832862-A2.  
 PD 30-JUL-1998.  
 PF 29-JAN-1998; E00484.  
 PR 29-JAN-1997; EP-101339.  
 PA (FIOH) FIOHE L.  
 PI FIOHE L, Hutter B, Kolk A, Singh M;  
 DR WPI: 98-427958/36.  
 PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum  
 PT - used for diagnosis of tuberculosis and other mycobacterial  
 PT diseases, also for treatment and prevention, for drug screening and  
 PT for bio-transformation  
 PS Disclosure: Page 11: 57pp; German.  
 CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated  
 CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis  
 CC and other mycobacterial infections (including 'swimmers' disease', caused  
 CC by M. marinum, a fish pathogen) in humans or animals. The protein can  
 CC also be used for control of epidemics and for vaccination, to screen for  
 CC agents with anti-mycobacterial activity, and in bio-transformations that  
 CC are specific for L-alanine. Also mycobacteria can be identified by  
 CC analysis of genomic ADH sequences. ADH is an antigen that is secreted  
 CC early during infection.  
 SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 98.5%; Score 1216; DB 1; Length 1260;  
 Best Local Similarity 99.3%; Pred. No. 1,26-244;  
 Matches 1232; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```

OY 1 ATCTTGACGATTAATGCACTTCTTCACTGAAAGCTACAGTATCGAGGGGTATC 60
DB 16 ATCTTGACGATTAATGCACTTCTTCACTGAAAGCTACAGTATCGAGGGGTATC 75
OY 61 ATGCGCGTGGTATTCGACGAGACCAAAACAGC-----AATTCGGGTGGCCATC 114
DB 76 ATGCGCGTGGTATTCGACGAGACCAAAACAGTATTCGAGGTGGCCATC 135
OY 115 ACCCGCGCGGCGTGGCGAACTAACCCGTGTGGCCATGAGGTGCTCATCGAGGACGT 174
DB 136 ACCCGCGCGGCGTGGCGAACTAACCCGTGTGGCCATGAGGTGCTCATCGAGGACGT 195
OY 175 GCGGAGAGGGGCTCGGCTATCAGCGAGCGGATTTCAAGGCGGACAGCGGCACTGCTC 234
DB 196 GCGGAGAGGGGCTCGGCTATCAGCGAGCGGATTTCAAGGCGGACAGCGGCACTGCTC 255
OY 235 GGCACCGCGGACAGGTGTGGCGGACGCTGATTTATGCTCAAGGTCAAGAACCGATA 294
DB 256 GGCACCGCGGACAGGTGTGGCGGACGCTGATTTATGCTCAAGGTCAAGAACCGATA 315
OY 295 GCGGCGGAATACGGCGCGCTGCGACAGCGGCGAGATCTTGTACGTTCTTCATTTGGCC 354
DB 316 GCGGCGGAATACGGCGCGCTGCGACAGCGGCGAGATCTTGTACGTTCTTCATTTGGCC 375
OY 355 GGGTACGCTGTGACGAGTGGCGGTGTGATTCGGGACACGATTCGCTACGAG 414
DB 376 GGGTACGCTGTGACGAGTGGCGGTGTGATTCGGGACACGATTCGCTACGAG 435
OY 415 ACCGTCACAGCCGCGACGCGGCGACTACCTGCTTGGCGGATGAGGAGGAGTGGCGGT 474
DB 436 ACCGTCACAGCCGCGACGCGGCGACTACCTGCTTGGCGGATGAGGAGGAGTGGCGGT 495
OY 475 CGACTGCGCCCGCAGGTTGGCGTTACCACTGATGCAACCAAGGGGCGCGGTGTG 534
DB 496 CGACTGCGCCCGCAGGTTGGCGTTACCACTGATGCAACCAAGGGGCGCGGTGTG 555
OY 535 CTGATGGGCGGCGGCGGCGGCGGAGAACGCGGCGAGCGTGTGATGCGGCGGCGGAC 594
DB 556 CTGATGGGCGGCGGCGGCGGCGGAGAACGCGGCGAGCGTGTGATGCGGCGGCGGAC 615

```

595 GCCGCTACACGACGACCCGATCGCCAAAGGATGGGCGGCGACCGTTACGGTTAGAC 654  
 616 GCGGGGTACAGACGACCCGATCGCCAAAGGATGGGCGGCGACCGTTACGGTTAGAC 675  
 655 ATCAACATGACAACTTGGGCACTCGACGCGGATTCGCGGCGGATCCACACTCGC 714  
 676 ATCAACATGACAACTTGGGCACTCGACGCGGATTCGCGGCGGATCCACACTCGC 735  
 715 TACTCATGCGCTACGAGCTCGAGGGTGGCTGAAGCTGCGACCTGGATGGGGCC 774  
 736 TACTCATGCGCTACGAGCTCGAGGGTGGCTGAAGCTGCGACCTGGATGGGGCC 795  
 775 GTCTGCTGCGACGCGGCAAGGACCCAAATTAAGTCTCGAATTCAGTTCGCGCATATG 834  
 796 GTCTGCTGCGACGCGGCAAGGACCCAAATTAAGTCTCGAATTCAGTTCGCGCATATG 855  
 835 AAACGAGTGGTACTGGTGGATATAGCATGACGAGGGGGGCTGTTTCAAGAGCTCA 894  
 856 AAACGAGTGGTACTGGTGGATATAGCATGACGAGGGGGGCTGTTTCAAGAGCTCA 915  
 895 CGACCGACGACCTACGACGACCCGACGTTGCGCGTGACGACAGCGTGTTCAGCGTG 954  
 916 CGACCGACGACCTACGACGACCCGACGTTGCGCGTGACGACAGCGTGTTCAGCGTG 975  
 955 GCGAATCATGCGCGCTGCGTGGTGGGAGAGCTCGACCTACGCGCTGACCAACGCGAGATG 1014  
 976 GCGAATCATGCGCGCTGCGTGGTGGGAGAGCTCGACCTACGCGCTGACCAACGCGAGATG 1035  
 1015 CCGATATGCTGCGAGCTTGGCGACGATGCTGGGCGGCGGCTGCGGATTCGCGCA 1074  
 1036 CCGATATGCTGCGAGCTTGGCGACGATGCTGGGCGGCGGCTGCGGATTCGCGCA 1095  
 1075 CTAGCCAAAGGCTTTTGCAGCGACGAGAGGGGCGTACTGTCGAAGCGGTGGCGACCGAC 1134  
 1096 CTAGCCAAAGGCTTTTGCAGCGACGAGAGGGGCGTACTGTCGAAGCGGTGGCGACCGAC 1155  
 1135 CTGGGGGTGGCGTTCACGAGCGCGGCGGAGGCTGCTGCGCTGACCTGCGGCGCTGTTAC 1194  
 1156 CTGGGGGTGGCGTTCACGAGCGCGGCGGAGGCTGCTGCGCTGACCTGCGGCGCTGTTAC 1215  
 1195 GCCGACGACACNTGCGAGTAAAGGAGCATGATGTCGAC 1235  
 1216 GCCGACGACACNTGCGAGTAAAGGAGCATGATGTCGAC 1256

**RESULT 3**  
 V49625 ID V49625 standard; DNA; 1260 BP.  
 AC V49625:  
 DT 20-NOV-1998 (first entry)  
 DE Mycobacterium tuberculosis Alamine dehydrogenase.  
 KW ss: Alamine dehydrogenase; tuberculosis; TB; mycobacteria; L-amine.  
 OS Mycobacterium tuberculosis.  
 PN MO9836089-A2.  
 PD 20-JAN-1998.  
 PR 28-JAN-1997: EP-101338.  
 PA (FLOH) FLOH L.  
 PI Flohe L, Hutter B, Kolk A, Singh M;  
 DR WCI; 98-457123/39.  
 PT Diagnosis of, e.g. tuberculosis from alamine dehydrogenase activity  
 PT useful for, e.g. for diagnosis, differentiation of strains,  
 PT monitoring vaccination and identification of mycobacterial  
 PT inhibitors  
 PS Claim 13: Fig 2.3: 55pp: German.  
 CC The Mycobacterium tuberculosis alamine dehydrogenase (AlaDH) is in  
 CC the production of kits for diagnosing tuberculosis (TB) and other  
 CC mycobacterial infections in humans or animals. Kits are used for direct  
 CC diagnosis of TB on clinical samples (e.g. body fluids) and can  
 CC differentiate between pathogenic and non-virulent strains, e.g. for  
 CC identifying pathogens of the M. tuberculosis (M. t.) complex. The kit may  
 CC also be used to identify substances that inhibit mycobacteria, for  
 CC combatting epidemics and for vaccination follow-up. Oligonucleotides

CC derived from AlaDH are used similarly in diagnostic hybridisation tests,  
 CC also for culture confirmation of isolated strains and for chromosome  
 CC fingerprinting to detect/differentiate between mycobacteria, and for  
 CC L-amine-specific biotransformation reactions. AlaDH is an early  
 CC antigen, present extracellularly after only a few days of growth, making  
 CC it an ideal drug target.  
 SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 98.5%; Score 1216; DB 1; Length 1260;  
 Best Local Similarity 99.3%; Pred No 1.2e-244;  
 Matches 1232; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

1 ATCTTGACATTAATGCACTTCTTCACTGAGCGCTACAGATGAGAGGGGTAATC 60  
 16 ATCTTGACATTAATGCACTTCTTCACTGAGCGCTACAGATGAGAGGGGTAATC 75  
 61 ATGCGGTGGTATTCCGACGAGCAAAACAAACG-----AATTGCGGTGGCCATC 114  
 76 ATGCGGTGGTATTCCGACGAGCAAAACAAACGAAATTCAAATTCGGGTGGCCATC 135  
 115 ACCCGGCGCGGCGTGGCGAATCAACCGCTGCGCATGAGTGTCTATCCAGGACAGT 174  
 136 ACCCGGCGCGGCGTGGCGAATCAACCGCTGCGCATGAGTGTCTATCCAGGACAGT 195  
 175 GCGGAGAGGCGTGGGCTATCAACGAGCGGATTTCAAGGCGGCGGCGCACTGGTC 234  
 196 GCGGAGAGGCGTGGGCTATCAACGAGCGGATTTCAAGGCGGCGGCGCACTGGTC 255  
 235 GGCACCGCGGACGAGTGGGCGGCGGCGGCTGATTTATGCTCAAGGTCAAAAGACGATA 294  
 256 GGCACCGCGGACGAGTGGGCGGCGGCGGCTGATTTATGCTCAAGGTCAAAAGACGATA 315  
 295 GCGGCGAATTAAGCGCGCGCTGCGACAGGCGAGATCTTGTTCAGTTCGATTTGGCC 354  
 316 GCGGCGAATTAAGCGCGCGCTGCGACAGGCGAGATCTTGTTCAGTTCGATTTGGCC 375  
 355 GCGTCACTGCTTGGACCGC GCGTGTGATTCGCGGACGACGATTCGCTTGGATTCG 414  
 376 GCGTCACTGCTTGGACCGCATGCTGTGATTCGCGGACGACGATTCGCTTGGATTCG 435  
 415 ACCGTCAACGCGCGGCGGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGG 474  
 436 ACCGTCAACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 495  
 475 CGACTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 534  
 496 CGACTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 555  
 535 CTGATGGGGGGGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 594  
 556 CTGATGGGGGGGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 615  
 595 GCCGCTACACGACGACCCGATCGCCAAAGGATGGGCGGCGGCGGCGGCGGCGGCGG 654  
 616 GCGGGGTACAGACGACCCGATCGCCAAAGGATGGGCGGCGGCGGCGGCGGCGGCGG 675  
 655 ATCAACATGACAACTTGGGCACTCGACGCGGATTCGCGGCGGATCCACACTCGC 714  
 676 ATCAACATGACAACTTGGGCACTCGACGCGGATTCGCGGCGGATCCACACTCGC 735  
 715 TACTCATGCGCTACGAGCTCGAGGGTGGCTGAAGCTGCGACCTGGATGGGGCC 774  
 736 TACTCATGCGCTACGAGCTCGAGGGTGGCTGAAGCTGCGACCTGGATGGGGCC 795  
 775 GTCTGCTGCGACGCGGCAAGGACCCAAATTAAGTCTCGAATTCAGTTCGCGCATATG 834  
 796 GTCTGCTGCGACGCGGCAAGGACCCAAATTAAGTCTCGAATTCAGTTCGCGCATATG 855  
 835 AAACGAGTGGTACTGGTGGATATAGCATGACGAGGGGGGCTGTTTCAAGAGCTCA 894  
 856 AAACGAGTGGTACTGGTGGATATAGCATGACGAGGGGGGCTGTTTCAAGAGCTCA 915





Matches 604; Conservative 0; Mismatches 458; Indels 12; Gaps 4;

```

OY 61 ATGCCGCTGCTATCCGACGACCAAAACAAAGATTCGGGTGCTATCCCGC 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 ATGAAGATCGGCTATCCAAAATAAATAAATAAATAAATAAATAAATAAATA 60
OY 121 GCGGCGCTGCGGACCTAACCCGCTGAGGTGCT---CATCAGGACAGTGC 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 GAGGCGCTGATGACCTGCTCAAAACGGGCGCTAGAGTGTATGTGACAGCAAG 120
OY 178 GAGAGG-CTGCGCTATACCGACCGGATTTCAAGCGCGGCGGCGCAATGTCGGC 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 GCGGCTGCGGCTGCTTCCGATTCGAGTGTGAAAAAGCGGCGGAGCTGATGCTGC 180
OY 238 ACCGCGACAGAGTGTGGGCGGACGCTATTTATGCTCAAGTCAAAAGACGATACG 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 CGAACGTGAGAGATGCTTGGACGCGGAGATGTTGAAAGTGAAGAGCGGCTGCT 240
OY 298 GCGGATACGCGCGCTGCGACACGGGACATCTTGTTCATGCTTGTGCTATTTGCGCG 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 CGAGAGTTCGGCTATTTGCGCGCGGATTTGTTTATGCTATTTGCTATTTACCGCG 300
OY 358 TCACGTGCTTGCACGATGCTGTTGATTCGCGACACGCTCAATTGCTACGAGACC 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 GCGGACGCGCTGACGAAAGCGCTGCTGACGAAAAAGTGTGCGGCTGCTTACGAGAC 360
OY 418 GTCCGACGCGCGGACGCGGACCTACCTGCTGCGCGGATGAGGACGCGGCTGGA 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 GTGCGCTGCGGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
OY 478 CTGCGCGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 421 ATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
OY 538 ATGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 481 CTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
OY 598 GCGCTGACGACGCGGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 541 GCGGACGACGCGGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
OY 658 AACATCGACAACTTTCGCAACTGACGCGGATTTCTGCGCGGCTGCTGCTGCTGCT 717
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 601 AACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
OY 718 TCATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 777
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 661 TCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
OY 778 CTGCTGCGGCGGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 837
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 721 TTGATCCGCGGCGGCGGAA---AGCGAAGCTGTGAGGAGAAAGATGCTGCTGCTGCT 777
OY 838 CCAGGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 894
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 778 CCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 837
OY 895 CGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 954
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 838 CGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 897
OY 955 GCGAAGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1014
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 898 GCGAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 954
OY 1015 CCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1074
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 955 CCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1014
OY 1075 CTAGCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1015 CTGTTAAAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1068

```

```

RESULT 6
ID V52155/c
V52155 standard; DNA; 28171 BP.
AC V52155;
DT 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:22.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
OS Streptococcus pneumoniae; vaccine; pharmaceutical composition; ds.
PN MO9818931-A2.
PD 07-MAY-1998.
PF 30-OCT-1997; U19588.
PR 31-OCT-1996; U5-029960.
PI (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA.
DR WPI: 98-272225/24.
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
PS Claim 1: Page 273-289; 1409bp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S. pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridize to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 28171 BP; 8507 A; 5080 C; 6440 G; 8144 T;

Query Match 17.2%; Score 212.4; DB 1; Length 28171;
Best Local Similarity 53.9%; Pred. No. 5,7e-36;
Matches 533; Conservative 0; Mismatches 436; Indels 19; Gaps 4;

OY 61 ATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2171 ATGTTATCGGAATCCCAAAAGAAATTAATAAGCAAAACCGTGCCTCACACT 21652
OY 121 GCGGCGGCGGGAACCTAACCCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21651 GCGAGTTCATGACTTATGTTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 21592
OY 181 GAGGCTGCGCTATACCGACGCGGATTTCAAGCGCGGCGGCGGCAACTGTGCGGACC 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21591 CTCGCTTCTGCTTACTGATGCTGACTCAAAAGCAAGAGCTGAGATGTGCTGACT 21532
OY 241 GCGGACGAGGTGCGGCGGCGGCTGATTTATGCTCAAGTCAAAAGCAAGAGGCGG 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21531 GCTGCTGTAAGCTTGGG---AGCAGAGTGTGTGAAATAAAGAAATCTTTAACTTCT 21475
OY 301 GAATACGCGCGCTGCGACACGCGGAGATTTGTTCAAGTCTTGTGATTTGCGGCTGCA 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21474 GAATACGCTTACTTGGCGGACGATCTTCTTCTTACCTTGTGACATAGCGCGCTGCT 21415

```





```

Db 9043 AAGATATTAGTATGATATGAGACTGTGAGTTACAGACGAGATCTTACATTGTGA 8984
Oy 451 GCCCGGATGAGGAGAGTCCCGGTGACACGCCGCCAGGTGGGGCTTACACCTGATG 510
Db 8983 TCACCAATGAGAGGTACAGAGAAAGTATAGCTCAAGTTGGGCAAGTTCCACAA 8924
Oy 511 CGAACCAAGGCGCGCGGTGTGATGGGGGTCCCGCGGTGACCGCGCGAC 570
Db 8923 AACTTAATGATGATGATGAGATTTACTAGTGGTCCAGAGTACTTAAAGGTAA 8864
Oy 571 GTCTGGTGTATGGCGCGCGGACCGCGCTACAAAGCAGCCCGCATGCGCAAGGCGATG 630
Db 8863 GTAACTATATGCGGTGGTGCAGAGAGAACAAATGCGCTAAATTCATAGAGACTA 8804
Oy 631 GCGCGACCGCTTACAGTTTACATCAATCAGCAAACTCGGCACTGACCGCGAG 690
Db 8803 GGTGAGATGTTACGATTTTAAATGTTTAAATCCAAAGCTTTACAAATTAATGATTTTA 8744
Oy 691 TTCTCGCGCGGATCCACACTGCTACTCATCGGCTTACAGCTCGAGGCGTCCGTCAAA 750
Db 8743 TTCTGGGAGACGTATACATTAATTAATGCAAAATCCGTTGAATATGATTTAATAA 8684
Oy 751 CGTGGCGACCTGTATTTGGGCGCTCCGTGGTCCAGGCGCAAGCGCAATTAATGTC 810
Db 8683 CAAAGTATTTAGTATAGTACAGTATTTAAATTCAGAGGTCTAAAGCGCGCAAGCTTGT 8624
Oy 811 TCGAATTCATCTGTGCGCGATGAAACAGAGTGGGCTAGTGGTATTTACCATCGAC 870
Db 8623 ACAGAGACATATTAACAAATGAAATAGGTGATATTAATTAATGATTTGAT 8564
Oy 871 CAGGCGCGCTGTTTGAAGGCTC---ACGACGACACCTTACAGACACCGACGTTCCG 927
Db 8563 CAAAGCGGATTTTGAAGCACTGATAAAATTAACAGACATGATGATCTCTATATAT 8504
Oy 928 GTGACAGACAGCTTTTACTGCTGTGGGAGACATGCCCGCTCGGTGGCGGAAGAGCTG 987
Db 8503 AAGCATGTGTGTTTCAATTAATGCAATATGCAAGTGCAGTACCGGCTACTTCG 8444
Oy 988 ACCTAGCGGCTGACCAAGCAGATGCGGTATGCTGCTGACGATTCGCGACCATGGCTGG 1047
Db 8443 ACGTTAGCTTAAATTAATGCTACGCTACCTTATGGCTCATGCTAGCTAATAAAGGTA 8384
Oy 1048 CGGCGCGCGTCCGCTCGAATCGGCACTAGCCAAAGCTCTTGCAGCGACGAGG 1103
Db 8383 AGAGAAGCATTTAAATCAATCAATCATATCATAGTTTAAATACCTTACAAAG 8328

RESULT 8
ID 167971 standard; DNA: 1074 BP.
AC 167971;
DT 15-JUL-1997 (first entry)
DE H. pylori membrane protein ORF 05cp20518orf61.
KW Vaccine; prevention; treatment; infection; identification;
binding compound; bacterium; life cycle; activator; bacteri-
inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
membrane; amino acid; metabolism; ds.
OS Helicobacter pylori.
FH Key
FT 1. 1074
FT cds
FT /*tag a
FT /note- 'no stop codon given'

MO9640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR) ASTRA AB.
PI Berglund OT, Smith D, Mellgaard BL.
DR WPI: 97-052306/05.
DR P-BSDB: W20718.
PT Helicobacter pylori nucleic acid sequences and related
polyptide(s) - useful for vaccines to treat or prevent H. pylori

```

```

PT Infection, and to detect Helicobacter
PS Claim 27; page 819; 1481p; English.
CC The present sequence encodes a Helicobacter pylori membrane
CC protein likely to contain four membrane spanning regions.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 35679) was determined from
CC overlapping contigs generated by mechanically shearing the
CC bacterial DNA. The sequences were analysed for ORF of at least 180
CC nucleotides, and the predicted coding regions defined by computer
CC evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences
CC of interest, particular regions can be isolated from H. pylori by
CC PCR amplification for recombinant polypeptide production, e.g. in
CC E. coli hosts.
SQ Sequence 1074 BP; 302 A; 189 C; 297 G; 286 T;

```

```

Query Match 7.78; Score 94.8; DB 1; Length 1074;
Best Local Similarity 46.18; Pred. No. 1,2e-11;
Matches 370; Conservative 0; Mismatches 417; Indels 15; Gaps 1;

```

```

Oy 265 GATTATTTGCTCAAGGTCAAGAACCGATAGCGGCGGATACGGCGCGCTCGACACGG 324
Db 133 GATTGTTGCTCAAGGTCAAGAACCGATAGCGGCGGATACGGCGCGCTCGACACGG 192
Oy 325 CAGATCTTGTTCACGCTTCTTCATTTGGCGCGCTGACGCTGACACGATCGCTTGTG 384
Db 193 GCGACTCTGTATTAGTATTGATTTAGCTATCAAAAGCTGTGTGAATTTTAT 252
Oy 385 GATTCGCGACACGCTCATTTGCTACGACACGCTGACAGCGCGGACGCTACCC 444
Db 253 AATAAATAATCATCTTATTTGCACTGAAACCATTTGGCGGCGCTAAAGACTACCT 312
Oy 445 CTGCTGCGCGGATGAGGAGTGGCGGCTGACGCGCGCGCGCGCGGCTGGG 496
Db 313 ATTTAGGCGCTATACGCTGTGTGGTGGAGGTTGGCTGCGCATTTAGACACATTAT 372
Oy 496 -----GCTTACCACTGATGCGAAGCCAGGCGCGCGGCTGTGCTGATGGCGGGTG 549
Db 373 TTACTGCTTTAGAGCATGTTAAAGGTTTATGGTAAAGGCGGTATCTAGGGGTTTG 432
Oy 550 CCGGCGTGCAGACCGCGCGACCTGTGTGATGCGCGCGCGCGCGCGCTTACAGCA 609
Db 433 TGGGTGCGCAAGGCGTAAATGCTGTATTTGAGGCGGTGTGGTGGATGGAAGC 492
Oy 610 GCCGCGATCGCCAAAGGCGCGCGCGCTTACGCTTACGATCAATCAATGACGACAA 669
Db 493 GCGAAAGCTTAAAGCAATGGGGCTTAAAGATTTTAAAGATTTAGACATGCTTAA 552
Oy 670 CTTGCGCACTGACGCGGATCTTGGCGCGGATTCACACACTGCTACTATCGGCTAC 729
Db 553 TTACAAACCAACCTTATTAATCATTTGATTTAGAGCTTAAAGGTGAATGAAGCC 612
Oy 730 GAGCTGAGGCTGCGCTCAAGCTGCGCATGCTGATTTGGCGCGCTGCTGCGGCG 789
Db 613 AATATCATTAAGGCTTAAAGGCGGCGGTGGGCTAGTGGAGGCGGTGCTGTACAGCG 672
Oy 790 GCGAAGGACCAATTTAGTCTGCAATTCATTTGCGGCAATTAAGAACCAAGGTCGGTA 849
Db 673 AGCCAAACCTTAAAGTATCTTAAAGGATTTAAATTAATGACGACAGCAAGGGTA 732
Oy 850 CTGTGATATAGCATGACAGGCGGCGGTGTTGAGGCGCTACGACGACGACCATC 909
Db 733 GTCAATGATGTGTGAGTTAGGGGCGGATTAAGGCGCATACGCAAGCAAGCAT 792
Oy 910 GACACCGGAGCTGCGCTGACAGACGCTGTTTACTGCTGGGCAAGACCGCGCC 969
Db 793 TCTAACCGGTGTATGTGGAAGAAAGTTTGTTCATATGCGCTGCGCAAGATGCCAGG 852

```

QY 1030 CTGCGACCATGGCTGGCGGG 1051

Db 913 TATTTAGCATGGCTGGAAGG 934

**RESULT 9**

ID	standard; DNA; 544 BP.
NC	W43039.

09-NOV-1998	(first entry)	DT
Streptococcus pneumoniae	polypeptide coding region.	DE
Polypeptide; ORF; open reading frame; Infection; Bacterial;		KM
Streptococcal; Bacteremia; diagnosis; prophylaxis; ds.		OS
Streptococcus pneumoniae.		KS
Location/Qualifiers		FT
complement (59. .334)		FT
/**tag- a		FT
/note- "polypeptide"		FT
W09823631.A1.		PN
04-JUN-1998.		PD
24-NOV-1997. U21976.		PF
27-NOV-1996: US-031879.		PR
(SMIR ) SMITHKLINE BEECHAM CORP.		PA
(SMIR ) SMITHKLINE BEECHAM PLC.		PA
Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO,		P1
Reid RH, Zarfos PN;		P1
WPI: 98-32654/28.		PI
P-PSDB: W62769.		DR
Streptococcus pneumoniae polynucleotides - useful for developing		PT
PT products for diagnosis, prevention and treatment of infections e.g.		PT
pneumonia, bacteremia, meningitis or endocarditis		PT
Claim 1: Page 165-166: 181pp: English.		CS
The sequence is that of a Streptococcal polypeptide coding region.		CC
The polypeptide can potentially be used for the diagnosis and		CC
prevention of bacterial infections, especially sp infection.		CC
It may be used for the treatment of diseases such as otitis media,		CC
conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleura		CC
empyema, endocarditis or infection of the cerebrospinal fluid.		CC
Sequence 544 BP: 149 A; 121 C; 124 G; 150 T;		Q0

Query Match	5.6%	Score 69.2;	DB 1;	Length 544;
Best Local Similarity	53.4%	Pred. No. 2.4e-06;		
Matches 191; Conservative	0;	Mismatches 163;	Indels 4;	Gaps 2

QY	691	TTTCGGCGCCGGATCCACACATCGCTACTCATCGGCGCTGCAGCTGCAGGGGTGGCGTAAA	750
Db	543	TTTGGAGTCAAAATTCAAATTCCTATATCTTAATCTAATTCATTCAACCATTTGAAGCAAGATGGA	488
QY	751	CGTGCCCAACCTGGTGGATTTGGGGCCGCTGCTGTGCGCAGGCGCCAAAGCACCCAA	810
Db	483	GATCTCTATGTGGGATTTGGAGCCATTTCTCACTCCCTGGTGCMAAAGCACCAGATTTGGTG	422
QY	811	TGCAATTCACCTCTGTCGGCGCATATCAAAACAGATCGCGTAC	870
Db	423	ACAGATGAGATGGTCAAAACAAATGCGTCCAGGCTGTGA	365
QY	871	CAGGCGCGCTGTTCGA	927
Db	364	CAAGGTGGCTTATCGGAACACAGCTGCAGCGTGTGCAACGCGACGATGAACCGTTATGAA	305
QY	928	GTGCGACGACGCTGTTTACTGTGCTGGGCGAACAATGGCCGCTGCGTGGCGAAMACGTGG	987
Db	304	AAACACGCTGTCTTC	245
QY	988	ACCTACGCGGTGACCAACGACGACATGGCGGATGTGCTCAGCTTGGCCGACATGGCT	1045
Db	244	ACCATCGCCTTACCCAAATGTACCTCTCTTATATGGAAGCTTTGGCTGGCAAAAGGAT	187

RESULT	10
V74370	
ID	V74370 standard; DNA; 31096 BP.
AC	V74370.

DE 10-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SEQ ID #59.  
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.  
 NC 000000

FT	key		Location/Qualifiers	
FT		/tag-	1201.	.1260
FT	misc_feature	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT		/tag-	3001.	.3060
FT	misc_feature	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT		/tag-	4801.	.4860
FT	misc_feature	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT		/tag-	6601.	.6660
FT	misc_feature	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT		/tag-	8401.	.8460
FT	misc_feature	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT		/tag-	10201.	.10260
FT	misc_feature	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT		/tag-	12001.	.12060
FT	misc_feature	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT		/tag-	13801.	.13860
FT	misc_feature	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT		/tag-	15601.	.15660
FT	misc_feature	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT		/tag-	17401.	.17460
FT	misc_feature	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT		/tag-	17401.	.17460
FT	misc_feature	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	

[illegible]

Query Match	4.78:	Score 57.6:	DB 1:	Length 31096:
Best Local Similarity	44.99:	Prod No. 0.00085:		
Matches 440:	Conservative 0:	Mismatches 519:	Indels 21:	Gaps
QY	78	GACGAGACCAAAAAACAAGATTCCGGGTGGCCATACCCCCGCGCGGATCCGCGAACT	137	
DB	10	GTCAAGAAATTAAACAAAGGTGAAGCGCTGTAGCTTGACACCCGAAATATGTGGTAAATT	69	
QY	138	AATCCGCTGGTGGCCATGAGGTGTCATCCACGCGAGGCGCGGAGAGGCTCGGCTATCAC	197	
DB	70	AATCGATGCACGACATTAAGTAATTTGTTGAAAAAATATGCTGGCATTGTTCAGGATTTTC	129	
QY	198	CGACGGGATTTCAAGGGCGGAGCGCGCAACTGTGTGGCACCAGCCGACAGGTGTGGC	257	
DB	130	TAAACATATGTATGTAAGAAAGAGGCGCTAA---GATCGTAATCAGCAAGCAAGCATGG--	185	
QY	258	CGACGCTATTATTATGTCAGAGTCAGAAAGACCGATAGAGCGCGGATATAGCGCCGCTGG	317	
DB	165	-GAGCTATCTTTGTTATCAAGTAAAGAAAGAACCTCATGAAGAGGATATCAATATTCAA	243	
QY	318	ACACGGGAGATCTTGTTACAGTTCTTGCAATTTGGCCGCGCTCACTGCTTGCACGATGC	377	
DB	244	AAAGATCAAAATTTATCTGGGATTTTATACCTAGCATCTTCAAAAGAAATAGTAGAAA	303	
QY	378	GTTGTGATTCGGGACACAGTCATATGCTCTACAGACCGTCCAAACCGCCGAGCGCC	437	
DB	304	AATGCAGAAATTGGTGTACTCTGAGTACTGTGTAACCAT--TATAAAAATGGAAA	360	
QY	438	ACTACCCCTGCTTCCCGGATGAGACGGAAGTCGCGCGTGCATCTGCCCGCCAGGTGGCGC	497	
DB	361	AGCAATATTATTAGCGCCCAATGATGCTTATAGCAGGTCAACGCTCAGCAATTATGGAGC	420	
QY	498	TTACACCTGATGGCAACCCAGAGGGGCGCGGTGTGCTATGAGGGGGGTGCCCGGCT	557	
DB	421	TTACTACTCTGAACACAACTGGTGTCAAGGTACTTTAGTACTGTGTACTATGAAAA	480	
QY	558	CGAACCGGCGCAGC-----TCGTGTGATATCGCGCGCGCACCGCCGCTACAAAGC	608	
DB	481	TGTGATATATCTGTGTAGTACATATGTGATTTTGGGTGGGAGTAGCAGCAAAATGTC	540	
QY	569	AGCCCGCATCGCCCAAGGCATGGGCGCGGACCGTTACGGTTCGTAGATCAATATGACAA	668	
DB	541	AGCAATATGTCTTGGGACTAAATGCTAAAGTAATCATATTAGGATTAAACGATGACCG	600	
QY	669	ACTTGGCACTGCACGCGCGAGTTCTCGCGCCCGCATCCACACTC--GCTACTCATCGC	725	
DB	601	CATTAAATATCTTGAGATATGTAATGTAATGTCAGAAAAAGATGTCAGTAGTCAATCAACAC	660	
QY	726	CTACAGCTCGAGGGTCCGCTCAAACTGCCGACCTGTGTATGAGGGCGCTCTGTGTGC	785	
DB	661	AGAAATTTTACACACAAATTAAGAAAGACAGATGTAATTTATTCTACAAATTTTAATTC	720	
QY	786	AGGCGCCCAAGCACCAATTAAGTCTGTGCAATTAATCTGTTCGCGCATATGAACACAGGTGC	845	
DB	721	AGGTCCGAAACCGCCAAATTTGGTTACTCTGTAGATGGTTAAATCAATGAAAAAGTTTC	780	
QY	846	GGTACTGCTGATATTAACCATTCAGCCAGGCGCGCTTTTGAAGGCTCACGACGACAC	905	
DB	781	AGTATTATCATATACCTATTAACCAAGGTGAGCATTTTGAACCAATTAATGACCAACTAC	840	
QY	906	CTACAGCACCCGACGTTCCCGCTGCACGACAGCTGTGTTTACTGCTGGCGACACATGCC	965	
DB	841	AATTTCTGATCTCACTGTATGAAAGAAAGAGGTGTGATTCTATTAATGAGTGTCCAAATCAAC	900	
QY	966	CGCCTCGGTGGCGAAGCGTCGACCTTAAGCGCTGACCAACGCGAGATGCCGTATGTGCT	1023	
DB	901	AGGAGCACTCCCAAGAACTTCAACCAATGGCATTAAGCACAAAGAAATTTGATTATATATT	960	
QY	1026	CGAGCTTGGCGACCATGCT 1045		
DB	961	AGAAATTTGTGACAAAGCT 980		



KW hybridisation; detection; vaccine; immunisation; infection; ss.  
 OS Mycobacterium sp.  
 PD MO9909186-A2.  
 PD 25-FEB-1999.  
 PF 14-AUG-1998; F01813.  
 PR 11-SEP-1997; FR-011325.  
 PR 14-AUG-1997; FR-010404.  
 PA (INSP) INST PASTEUR.  
 PI Giequel B, Llm EM, Pellicic V, Portnoi D, Goguet de la Salmoniere Y,  
 PI Guigueno A;  
 DR WPI: 99-181045/15.  
 DR P-PSDB: Y04952.  
 PT Mycobacterial DNA vectors containing reporter constructs - for  
 PT identifying coding or promoter sequences involved in  
 PT infection-associated protein expression  
 PS Claim 22; Fig 41S; 309pp; French.  
 CC Sequences X34001-X34252 represent nucleic acids encoding secreted  
 CC proteins from various Mycobacterium species microorganisms. The  
 CC nucleotide sequences can be used as primers and probes for methods  
 CC for detecting and identifying mycobacteria, especially belonging to  
 CC the M. tuberculosis complex. The encoded proteins can be used in  
 CC vaccines for immunisation against a bacterial or viral infection.  
 SQ Sequence 1413 BP; 187 A; 459 C; 540 G; 227 T;

Query Match 4.1%; Score 50.4; DB 1; Length 1413;  
 Best Local Similarity 43.0%; Pred. No. 0.021;  
 Matches 246; Conservative 0; Mismatches 326; Indels 0; Gaps 0;

QY 69 CGGATATCCGACGAGACCAAAACAGATTCCGGGTCATCCACCCCGCGCGT 128  
 DB 69 CGGATATCCGACGAGACCAAAACAGATTCCGGGTCATCCACCCCGCGCGT 128  
 QY 129 CGCGAATACCCCGTGTGGCCATGAGTCTCATCCAGCGAGTGCAGAGGCTC 188  
 DB 129 GCGCGGTTGTGGCGGCGGTCAGATTGTCGTCGCGATGATGGGTGGAGTGGC 188  
 QY 189 GCGTATCAGCAGCGGATTTCAAGCGCGACGCGACATGTCGCGACCGCGACA 248  
 DB 189 GCGAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 248  
 QY 249 GGTGTGGCGGACGCTGATTTATGCTCAAGGTCAAGAACAGATAGCGGGAATACGG 308  
 DB 249 GGGGTGGCGGAGTGGGTATACCGCGCGCGTACCGCGCGTACCGCGCTGGAAGAC 308  
 QY 309 CGCGCTGCGACACGCGCATTTGTCACGTTCTTGATTTGGCGGTCACGCTTG 368  
 DB 309 CGCGGACGCTGCGCATTTGTCACGTTCTTGATTTGGCGGTCACGCTTG 368  
 QY 369 CACGATGCGTGTGATTTCCGACGACGATTCATTTGCTAGAGACCGTCCACACCG 428  
 DB 369 GGTGTGCTGCTGAGAGCGCTTGCCTCCACATTCGCGGCGAGCACTGCGCGCGT 428  
 QY 429 CGAGCGGACATACCCCTGCTTCCCGATGAGCGAAGTCCGCGGTCACGCGCGCA 488  
 DB 429 CTACGACACCTTCGCGCGCGCTTCCCGACGACGCGTACCAACGTCGCGCGCGCA 488  
 QY 489 GGTGTGGCGGCTTACGCTGATGAGCAACCGAGGCGCGCGTGTGATGGGCGGT 548  
 DB 489 ATTGAGAGTATGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 548  
 QY 549 GCGCGGCTGAGACCGCGCGGATGATGATGATGATGATGATGATGATGATGATGAT 608  
 DB 549 TCGACTCAAGGAATTGGGTGATGATGATGATGATGATGATGATGATGATGATGAT 608  
 QY 609 AGCGCGCATCGCAACGCGATGGCGCGACCG 640  
 DB 609 CGTACAGCTCCACGACGACGCGGTCGCG 640

RESULT 14  
 X34206  
 ID X34206 standard; DNA: 1722 BP.

AC X34206: (first entry)  
 DT 06-JUL-1999  
 DE Mycobacterium species nucleic acid sequence 41T.  
 KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;  
 KW hybridisation; detection; vaccine; immunisation; infection; ss.  
 OS Mycobacterium sp.  
 PD MO9909186-A2.  
 PD 25-FEB-1999.  
 PR 14-AUG-1998; F01813.  
 PR 11-SEP-1997; FR-011325.  
 PR 14-AUG-1997; FR-010404.  
 PA (INSP) INST PASTEUR.  
 PI Giequel B, Llm EM, Pellicic V, Portnoi D, Goguet de la Salmoniere Y,  
 PI Guigueno A;  
 DR WPI: 99-181045/15.  
 DR P-PSDB: Y04953.  
 PT Mycobacterial DNA vectors containing reporter constructs - for  
 PT identifying coding or promoter sequences involved in  
 PT infection-associated protein expression  
 PS Claim 22; Fig 41T; 309pp; French.  
 CC Sequences X34001-X34252 represent nucleic acids encoding secreted  
 CC proteins from various Mycobacterium species microorganisms. The  
 CC nucleotide sequences can be used as primers and probes for methods  
 CC for detecting and identifying mycobacteria, especially belonging to  
 CC the M. tuberculosis complex. The encoded proteins can be used in  
 CC vaccines for immunisation against a bacterial or viral infection.  
 SQ Sequence 1722 BP; 231 A; 548 C; 657 G; 286 T;

Query Match 4.1%; Score 50.4; DB 1; Length 1722;  
 Best Local Similarity 43.0%; Pred. No. 0.021;  
 Matches 246; Conservative 0; Mismatches 326; Indels 0; Gaps 0;

QY 69 CGGATATCCGACGAGACCAAAACAGATTCCGGGTCATCCACCCCGCGCGT 128  
 DB 378 GCGACATGCGCGCGCGCTTGTGGCGGATATCCGGCGGTGACGCGCAACCCCTCGG 437  
 QY 129 CGCGAATACCCCGTGTGGCCATGAGTCTCATCCAGCGAGTGCAGAGGCTC 188  
 DB 438 GCGCGGTTGTGGCGGCGGTCAGATTGTCGCTCATGAGTGGTGGAGTGGC 437  
 QY 189 GCGTATCAGCAGCGGATTTCAAGCGCGACGCGCGACATGTCGCGACCGCGACA 248  
 DB 498 GCGAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 497  
 QY 249 GGTGTGGCGGACGCTGATTTATGCTCAAGGTCAAGAACAGATAGCGGGAATACGG 308  
 DB 558 GGGGTGGCGGAGTGGGTATACCGCGCGCGTACCGCGCGTACCGCGCTGGAAGAC 517  
 QY 309 CGCGCTGCGACACGCGCATTTGTCACGTTCTTGATTTGGCGGTCACGCTTG 368  
 DB 618 CGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 617  
 QY 369 CACGATGCGTGTGATTTCCGACGACGATTCATTTGCTAGAGACCGTCCACACCG 428  
 DB 678 GGTGTGCTGCTGAGAGCGCTTGCCTCCACATTCGCGGCGAGCACTGCGCGCGT 737  
 QY 429 CGAGCGGACATACCCCTGCTTCCCGATGAGCGAAGTCCGCGGTCACGCGCGCA 488  
 DB 738 CTACGACACCTTCGCGCGCGCTTCCCGACGACGCGTACCAACGCGCGCGCGCA 797  
 QY 489 GGTGTGGCGGCTTACGCTGATGAGCAACCGAGGCGCGCGTGTGATGGGCGGT 548  
 DB 798 ATTGAGAGTATGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 548  
 QY 549 GCGCGGCTGAGACCGCGCGGATGATGATGATGATGATGATGATGATGATGATGAT 608  
 DB 858 TCGACTCAAGGAATTGGGTGATGATGATGATGATGATGATGATGATGATGATGAT 608  
 QY 609 AGCGCGCATCGCAACGCGATGGCGCGACCG 640  
 DB 918 CGTACAGCTCCACGACGACGCGGTCGCG 949





Fri Jun 23 09:31:28 2000

us-09-362-485-3.rni

Page 1

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:01:15 ; Search time 341.15 Seconds  
(without alignments)  
470.560 Million cell updates/sec

Title: US-09-362-485-3  
Sequence: 1 ATCTTCAGATTATGAC.....ANGAGAGCATGATGTCGNC 1235

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47.6	3.9	1620	US-08-461-775-10	Sequence 10, Appl
2	47.6	3.9	28958	US-08-258-261B-6	Sequence 6, Appl
3	47.6	3.9	28958	US-08-456-837-6	Sequence 6, Appl
4	47.6	3.9	28958	US-08-457-342-6	Sequence 6, Appl
5	47.6	3.9	28958	US-08-457-646A-6	Sequence 6, Appl
6	47.6	3.9	28958	US-08-458-076A-6	Sequence 6, Appl
7	47.6	3.9	28958	US-08-764-233A-4	Sequence 4, Appl
8	47.6	3.9	28958	US-08-457-335A-6	Sequence 6, Appl
9	47.6	3.9	28958	US-08-729-214-6	Sequence 6, Appl
10	47.6	3.9	49377	US-08-764-233A-1	Sequence 1, Appl
11	47.6	3.7	2582	US-08-816-105A-2	Sequence 2, Appl
12	45.4	3.7	833	US-08-403-852D-3	Sequence 3, Appl
13	45.4	3.7	5392	US-08-403-852D-1	Sequence 1, Appl
14	42.4	3.4	1833	US-08-403-852D-6	Sequence 6, Appl
15	42.4	3.4	1833	US-08-173-508-3	Sequence 3, Appl
16	42.4	3.4	2185	US-08-265-310-3	Sequence 3, Appl
17	41.2	3.3	2185	US-08-461-775-11	Sequence 11, Appl
18	41.2	3.3	2668	US-07-642-734C-3	Sequence 3, Appl
19	41.2	3.3	20235	US-07-642-734C-7	Sequence 7, Appl
20	40.8	3.3	44377	US-08-804-227C-1	Sequence 1, Appl
21	40.8	3.3	44377	US-08-804-198-1	Sequence 1, Appl
22	40.6	3.3	4257	US-08-690-473-1	Sequence 1, Appl
23	40.6	3.3	12001	US-08-458-568A-11	Sequence 11, Appl
24	40.2	3.3	2064	US-08-343-428-1	Sequence 1, Appl
25	40.2	3.3	11219	US-07-642-734C-1	Sequence 1, Appl
26	39.6	3.2	474	US-08-403-852D-14	Sequence 14, Appl
27	39.6	3.2	1524	US-08-461-775-10	Patent No. 5512669

28	39.6	3.2	1525	7	5229279-1	Patent No. 5229279
29	39.4	3.2	459	4	US-08-387-942C-35	Sequence 35, Appl
30	39.4	3.2	1998	7	5122968-8	Patent No. 5122968
31	39.4	3.2	12588	4	US-08-387-942C-1	Sequence 1, Appl
32	39.2	3.2	420	1	US-08-470-179-148	Sequence 148, App
33	39.2	3.2	2353	6	PCT-US92-06840-1	Sequence 1, Appl
34	39.2	3.2	43280	3	US-08-804-227C-1	Sequence 1, Appl
35	38.8	3.1	8051	4	US-08-576-656A-2	Sequence 2, Appl
36	38.8	3.1	11219	2	US-07-642-734C-1	Sequence 1, Appl
37	38.6	3.1	1187	2	US-08-440-856A-2	Sequence 2, Appl
38	38.6	3.1	3231	2	US-08-074-121-4	Sequence 4, Appl
39	38.6	3.1	3231	6	PCT-US94-06447-4	Sequence 4, Appl
40	38.4	3.1	1215	4	US-08-947-726A-1	Sequence 1, Appl
41	38.4	3.1	2048	5	US-08-776-251-1	Sequence 1, Appl
42	38.2	3.1	2109	4	US-08-555-568B-20	Sequence 20, Appl
43	38	3.1	30001	1	US-08-125-468-1	Sequence 1, Appl
44	38	3.1	30001	3	US-08-474-933-1	Sequence 1, Appl
45	37.8	3.1	2588	3	US-08-796-414B-6	Sequence 6, Appl

#### ALIGNMENTS

RESULT 1  
US-08-461-775-10  
Sequence 10, Application US/08461775  
Patent No. 5858773  
GENERAL INFORMATION:  
APPLICANT: MAZODIER, Philippe  
APPLICANT: GUGLIEMTI, Gerard  
TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE  
TITLE OF INVENTION: INITIATION OF TRANSCRIPTION  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22131-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,775  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/050,313  
FILING DATE: 10-MAY-1993  
APPLICATION NUMBER: FR 9011186  
FILING DATE: 10-SEP-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Reuty, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 010830-035  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-2021  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1620 base pairs  
TYPE: nucleic acid  
STRANDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1620  
US-08-461-775-10



Db 25786 GCGTTCGTCAGACGCTCTCTCCCTGAGACGCGCCGCGCATCCGCCCTGCGCAGC 25845  
QY 685 GCGGAGTTCTGCGCGCGGATCCACACTGCTACTCATCGGCTAGAGCTCGAGGGTGC 744  
Db 25846 AAGCGCTACACACCGTCCGCGGAGGAGGCGCATGGCCCGCTCCAGCTCGCGCTCC 25905  
QY 745 GTCAAAAGTCCGACCTGTGATTTGGGCGCTCTGCTGCGCAGCGCC 792  
Db 25906 GACCTCCAGACTACCTGCTCCCTGGGCGAGCAGGCTCTCATCGCC 25953

## RESULT 3

US-08-456-837-6  
; Sequence 6, Application US/08456837  
; Patent No. 5643774  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James M.  
; APPLICANT: Beck, James Joseph  
; APPLICANT: Hill, Dwight Steven  
; APPLICANT: Ryals, John Andrew  
; APPLICANT: Gaffney, Thomas Deane  
; APPLICANT: Lam, Stephen Ting  
; APPLICANT: Hammer, Phillip E.  
; APPLICANT: Uknes, Scott Joseph  
; TITLE OF INVENTION: Genes for the synthesis of  
; TITLE OF INVENTION: antipathogenic substances  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ciba-Geigy Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/456,837  
; FILING DATE: 01-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/457,205  
; FILING DATE: 01-JUN-1995  
; APPLICATION NUMBER: 08/258,261  
; FILING DATE: 08-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rimer, James Scott  
; REGISTRATION NUMBER: 36,129  
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8614  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28958 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-456-837-6

Query Match 3.9%; Score 47.6; DB 1; Length 28958;  
Best Local Similarity 44.1%; Pred. No. 0.01;  
Matches 286; Conservative 0; Mismatches 37; Indels 3; Gaps 2;

QY 148 GGCATGAGTGTCTATCC/ GCAGTGGCGAGAGGGCTCGCTATACCGAGCGGAT 207  
Db 25306 GCGCAGAGACCTACACCTCGCGGATGTGGCTATTCGCTGACACCGCGCCAC 25385  
QY 208 TTCAGAGCGGAGCGCGGCACTGTGTGGCAGCGGCAAGTGTGGGCGCAGCTGAT 267  
Db 25366 TTCAGAGCGGCGCGCTGTGTAGCCCAACCGCGAGAGACTCTCTCGCGGCTGAC 25425  
QY 268 TTATGCTCAGAGTCAAAACAGATAGGGGGAAATAGCGCGCTGTGACAGGGGAC 327  
Db 25426 TCGCTGCGCAGAGACAGCGCGCGCGGAGACAGCTGTCTGAGGAGGAGGAGCGC 25485  
QY 328 ATCTGTTACGCTTTTCATTTGGCGGCGTCACTGCTTGACCGATGCTTTGGAT 387  
Db 25486 AAGCTGTCTTCTGCTCTTCTGCGCAAGCTGTGCAAGTGGAGGATAGCCCTCTCGTG 25545  
QY 388 TCGGCGACAGCTCATTTTC/ TACGAGACCTTCCAGACCGCGGCGGCGCAGCTACCCCTG 447  
Db 25546 CTGCACTCTCTCGCGCTCTTCGCGCTCAGCTGGAAGCATGCGAGCGCGCTGCTCT 25605  
QY 448 CTTGCGCGGATGAGGAGAGTCCGCGGTGACTCGCGCGGCGGAGTTGGGCTTACGAC -C 505  
Db 25606 CAGTGTGATGAGAGCTGTGCTGCTGCGCGCGGCGGAGGAGGCGCGCTCTCTGAC 25665  
QY 506 TCTGCGAACCAGAGGCGCGGCTGTGCTGATGAGGCGGCGGCGGCGGCTGCAACCGG 565  
Db 25666 CCGTGTGAGCTGTGACAGCGCGCTCTTGGCGGTGATGTCTCTGCGGCGGCTCTGG 25725  
QY 566 CCGAGC-TGCTGTGATGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 624  
Db 25726 CCGTGTGCTGCGGCTGAGAGCGCGCGCGCTGTGCGGCGGCGGCGGCGGCGGCGG 25785  
QY 625 GGCATGAGGCGGCGGCTTACGTTCTAGACATCAATGACAACTTCGCGCACTGAC 684  
Db 25786 GCGTTCGTGAGGCGGCTGTCTCTGCGGAGGAGGCGGCGGCGGCGGCGGCGGCGG 25845  
QY 685 GCGGAGTTCTGCGCGCGGATCCACACTGCTACTCATGCGCTAGAGCTCGAGGTTGCC 744  
Db 25846 AAGCGCTACACACCGTCCGCGGAGGAGGCGCATGGCGCGCTGACCTCGCGGCTCC 25905  
QY 745 GTCAAGCGTCCGACCTGTGATTTGGGCGCTCTGCTGCGCAGCGCC 792  
Db 25906 GACCTCCAGACTACCTGCTCCCTGGGCGAGCAGGCTCTCATCGCC 25953

## RESULT 4

US-08-457-342-6  
; Sequence 6, Application US/08457342  
; Patent No. 5652898  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James M.  
; APPLICANT: Beck, James Joseph  
; APPLICANT: Hill, Dwight Steven  
; APPLICANT: Ryals, John Andrew  
; APPLICANT: Gaffney, Thomas Deane  
; APPLICANT: Lam, Stephen Ting  
; APPLICANT: Hammer, Phillip E.  
; APPLICANT: Uknes, Scott Joseph  
; TITLE OF INVENTION: Genes for the synthesis of  
; TITLE OF INVENTION: antipathogenic substances  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ciba-Geigy Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/457,342  
APPLICATION NUMBER: US/08/457,342  
FILING DATE: 01-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/457,205  
FILING DATE: 01-JUN-1995  
APPLICATION NUMBER: 08/258,261  
FILING DATE: 08-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28958 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-457-342-6

Query Match 3.9%; Score 47.6; DB 1: Length 28958;  
Best Local Similarity 44.1%; Pred. No. 0.01;

Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

QY 148 GCCCATGAGTCTCTATCCAGGAGTGGCGGAGGCGTCCGATATCCAGCGGAT 207  
DB 25306 GCCCAGCAGACCTCACTCCGCGATGCGCTATTCGCTGCCACACCCGCGCCAC 25365  
QY 208 TTCAGGCGGCGAGCGCGCACTGTGCGACCGCGACAGGTGTGGCGCGACGCTAT 267  
DB 25366 TTCGAGCAGCGCGCGCTCTCGTAGCCCAACCGGAGAGACTCTCTCGCGCGTCCAG 25425  
QY 268 TTATTGCTCAGGTCAAGAACCGATAGCGGGGAATACGGCGCGCTCGGACAGGGGCG 327  
DB 25426 TCGCTCGCCCAAGAGCCCGCCGAGACCGTCTCTGAGAGCGAAGCCACGCGC 25485  
QY 328 ACTTGTTCAGCTCTCTGATTTGGCGGCGTCAAGCGTTCAGCGATCGTTGTTGAT 387  
DB 25486 AAGCTGCTTCTCTTCTTCTGCGCAAGGCTGCAAGTGGAAAGGATGCGCTCTGCTG 25545  
QY 388 TCCGACACCACTCAATTCCTACGAGACCGTCCAGACCGCGCGACGCTACCCCTG 447  
DB 25546 CTCGACTCTCTCGCGCTCTTCGCGCTCAGCTCGAAGCATGGAAGGCGCGCTCGTCT 25605  
QY 448 CTGCGCGCATAGCAGAGTCCGCGGTGCGACTCGCGCGCGCGCGGATTGGCGCTTAC 505  
DB 25606 CACGTGAGTGGAGCTCTGCGCGCTCTGCGCGCGCGAGAGGCGCGCGCTCTCGAC 25665  
QY 506 TGATGGAACCAAGGCGCGCGGTGCTGTATGGCGGGGCGCGCGCGTCCGAAACGG 565  
DB 25666 CGCGTGGAGCTGATCCCGCGCGCTTTCGCGCTATGTTTCCTGCGCGCGCTCTG 25725  
QY 566 CCGAGC-TGCTGATTCGCGCGCGCGCGCGCGCTACCAAGACCGCGCATTCGCGCA 624  
DB 25726 CGCTGCTGGGTGAGAGCGCGCGCGCGCTGTCGCGCAAGTCAAGGCGCGCGCGC 25785  
QY 625 GCGATGGCGCGCGCTTACGCTTCTAGACATCAACATGACCAACTTGGCAACTCGAC 684  
DB 25786 GCGTCTGTCGAGCGCTCTCTCTGAGGAGCGCGCGCGCTCGCGCTGCGCGAC 25845  
QY 685 GCGGATTTGGGCGCGCGATCCACTGCTACTGAGGCTGAGGAGTCCGAGGCTGCG 744  
DB 25846 AAGCGCTACCAAGCTCGCGCGCGCGCGCGCGCGCGCGCTGCGAGTCTGCGCGCTCC 25905

QY 745 GTCAAGTCCGACCTGATGTTGGGCGGCTCTGTTGCCAGCGCGC 792  
DB 25906 GACCTCCAGACTACTGCTCTCCCTGGGCGAGCGCTTCATCGCC 25953

## RESULT 5

US-08-457-646A-6  
Sequence 6, Application US/08457646A  
Patent No. 5679560  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James M.  
APPLICANT: Beck, James Joseph  
APPLICANT: Hill, Dwight Steven  
APPLICANT: Ryals, John Andrew  
APPLICANT: Gathney, Thomas Deane  
APPLICANT: Lam, Stephen Ting  
APPLICANT: Hammer, Phillip E.  
APPLICANT: Utnes, Scott Joseph  
TITLE OF INVENTION: Genes for the synthesis of  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: 93/08/457,646A  
FILING DATE: 01-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/457,205  
FILING DATE: 01-JUN-1995  
APPLICATION NUMBER: 08/258,261  
FILING DATE: 08-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28958 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-457-646A-6

Query Match 3.9%; Score 47.6; DB 1: Length 28958;  
Best Local Similarity 44.1%; Pred. No. 0.01;

Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

QY 148 GCCCATGAGTCTCTATCCAGGAGTGGCGGAGGCGTCCGATATCCAGCGGAT 207  
DB 25306 GCCCAGCAGACCTCACTCCGCGATGCGCTATTCGCTGCCACACCCGCGCCAC 25365  
QY 208 TTCAGGCGGCGAGCGCGCACTGTGCGACCGCGCGACAGGTGTGGCGCGAGCTGAT 267  
DB 25366 TTCGAGCAGCGCGCGCTCTCTGATGCCCAACCGCGAGAGCTCTCTCGCGCTGAC 25425



Sequence 4, Application US/08764233A  
Patent No. 5716849  
GENERAL INFORMATION:  
APPLICANT: Ligon, James M.  
APPLICANT: Schupp, Thomas  
APPLICANT: Beck, James J.  
APPLICANT: Hill, Dwight S.  
APPLICANT: Neff, Suezanna  
APPLICANT: Ryals, John A.  
TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 520 White Plains Road, P.O. Box 2005  
CITY: Tarrytown  
STATE: NY  
COUNTRY: USA  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764,233A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/729,214  
FILING DATE: 09-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/258,261  
FILING DATE: 08-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: 1506/CIP6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 541-8587  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28958 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORGANISM: Sorangium cellulosum  
IMMEDIATE SOURCE:  
CLONE: P98/1  
US-08-764-233A-4

Query Match 3.9%; Score 47.6; DB 2; Length 28958;  
Best Local Similarity 44.1%; Pred. No. 0.01;  
Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

QY 148 GCCCATGAGTCTCTATCAGCAGGATGCCGAGAGAGGCTGCTATCAGCAGCGGAT 207  
DB 25306 GCCCAGAGACTCACCCTCGCGATGTGGCTTATTCCTGCGCCACACCGCGCCAC 25345  
QY 208 TTCAGAGCGGCGAGCGGCACTGTGCGCACCGCGCACAGGTGTGGCGCGACCTGAT 267  
DB 25366 TTCGAGACCGCGCGCTCTAGCCCAACCGCGAGAGACTCTCTCCCGCTCGAC 25425  
QY 268 TTATGCTCAAGGTCAAGAACCGATAGCGGGGGAATAGCGCGCTCGGACAGCGGAC 327  
DB 25426 TCGCTGCGCCAGACAGCGCGCGCGCGAGCACCGTCTCGAGGAGCGGAGCGACGC 25485  
QY 328 ATCTGTACGCTCTGCTGCAATTTGGCCGCGCTGCTGCTTGCACCGATGCGTTGTGAT 387  
DB 25486 AAGCTGCTTGTCTTTCCTCGGCAAGGCTCGCAGTGGGAAGGATGCGCTCTCGCTG 25545

QY 388 TCCGACACCGATCAATTGCTTACAGAGACGTCACAGACCGCGGACGCTACCCCTG 447  
DB 25546 CTCGACTCTTCGCCCTGTTCCAGGCTCAGCTCAGAGATCGAGCGGCTGCTCT 25605  
QY 448 CTTCGCCGATGAGCGAAGTCCGCGGTGATCGCGCGCCAGTTGGGCTTACAC -C 505  
DB 25606 CACGTGAGTGAAGCTCTGCGCGCTCTGCGCGCGAGAGGCGCGCCCTCCCTCGAC 25665  
QY 506 TGATCCGAACCCAGAGGCGCGGCTGTCTGATGAGGCGGCTCCCGCGCTGACCGG 565  
DB 25666 CGCTGACGCTGTACACCGCGCTCTTTCGCTGATGCTTCTGCGCGGCTCTG 25725  
QY 566 CCGAG-TGCTGTGATGAGCGCGGACCGCGCTGCTGACAGCAGCGCCGATCGCAAC 624  
DB 25726 CGCTGCTCGGGGTAGAGCGCGCGCGCTCTGCTGCGGACATGAGGCGAGATGCGCC 25785  
QY 625 GGCATGCGCGGACCGCTTACGCTTACATCAATCAATGACAACTTGGCAACTGAC 684  
DB 25786 GCCTTCGTCGAGCGGCTCTCTCCCTGAGAGCGCGCGCATGCGCGCTCGCAGC 25845  
QY 685 GCGAGTCTGCGCGCGGATCCACACTGCTACTCATCGGCTTCAGAGCGAGGTGCC 744  
DB 25846 AAGGCTCACCACCTGCGCGGACAGGCGCATGCGCGCTGAGCTGCGCGCTCC 25905  
QY 745 GTCAACGTCGCGACCTGTGATTTGGGCGCTCTCTGTGTCAGCGCGCC 792  
DB 25906 GACCTCCAGACCTACTGCTCTCGGGGCGACAGGCTCTCATCGCC 25953

## RESULT 8

US-08-457-335A-6

Sequence 6, Application US/08457335A

Patent No. 5723759

GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Ligon, James M.

APPLICANT: Beck, James Joseph

APPLICANT: Hill, Dwight Steven

APPLICANT: Ryals, John Andrew

APPLICANT: Gaffney, Thomas Deane

APPLICANT: Lam, Stephen Ting

APPLICANT: Hammer, Phillip E.

APPLICANT: Ukens, Scott Joseph

TITLE OF INVENTION: Genes for the synthesis of

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba-Geigy Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/457,335A

FILING DATE: 01-JUN-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/457,205

FILING DATE: 01-JUN-1995

APPLICATION NUMBER: 08/258,261

FILING DATE: 08-Jun-1994

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129

REFERENCE/DOCKET NUMBER: CGC 1506/CIP3

TELECOMMUNICATION INFORMATION:





Db 25726 CGCTGCTGCGCGTAGAGCCCGCGCTGCTGCGCCACAGTCAGGCGAGATGCCGCC 25785  
QY 625 GGCATGGGCGCGCGCTTAGAGCTTACATCAACATCGACAAACTTCGGCAACTGAC 684  
Db 25786 GCCTTCGTCGAGCGCTCTCTCTCCATCGAGAGCGCGCCGCGCTTCGCGAC 25845  
QY 685 GCGGAGTTCGCGCGCGAGTCACACTGCTATCGGCTTACGAGTCGAGGTCGCC 744  
Db 25846 AAGCGCTACACCGTCGCGCGGAGCATGCGCGCGCTGCGAGTCGCGCGCTCC 25905  
QY 745 GTCAAGTGGCGCGCTGATTCGCGCGCTGCTGCGAGCGCGC 792  
Db 25906 GACCTCCAGACTACCTGCTCCTCGGCGAGCGCTTCATCGCC 25953  
RESULT 10  
US-08-764-233A-1  
Sequence 1, Application US/08764233A  
Patent No. 5716849  
GENERAL INFORMATION:  
APPLICANT: Ligon, James M.  
APPLICANT: Schupp, Thomas  
APPLICANT: Beck, James J.  
APPLICANT: Hill, Dwight S.  
APPLICANT: Neff, Snezana  
APPLICANT: Ryals, John A.  
TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 520 White Plains Road, P.O. Box 2005  
City: Tarrytown  
STATE: NY  
COUNTRY: USA  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764,233A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/729,214  
FILING DATE: 09-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/258,261  
FILING DATE: 08-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: 1506/CIP6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 541-8587  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 49377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Sorangium cellulosum  
IMMEDIATE SOURCE:  
CLONE: p98/1, pUL3, and pYKM15  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 383..760  
OTHER INFORMATION: /product- "Sorb"  
OTHER INFORMATION: /note- "This gene encodes a protein that is highly homologous

OTHER INFORMATION: the reductase domains of type I PKs such as eryA from  
OTHER INFORMATION: Saccharopolyspora erythraea."  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 927..19874  
OTHER INFORMATION: /product- "Sora"  
OTHER INFORMATION: /note- "Gene product is highly homologous to type I PKs t  
OTHER INFORMATION: are known to be involved in the synthesis of polyketide  
OTHER INFORMATION: compounds."  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 942..7115  
OTHER INFORMATION: /product- "Module 1 of Sora"  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 7203..12884  
OTHER INFORMATION: /product- "Module 2 of Sora"  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 13455..19616  
OTHER INFORMATION: /product- "Module 3 of Sora"  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 19871..46318  
OTHER INFORMATION: /product- "Sorb"  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 19870..24556  
OTHER INFORMATION: /product- "Module 1 of SorB"  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 24638..30820  
OTHER INFORMATION: /product- "Module 2 of SorB"  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 30881..35446  
OTHER INFORMATION: /product- "Module 3 of SorB"  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 35528..40114  
OTHER INFORMATION: /product- "Module 4 of SorB"  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 40190..46318  
OTHER INFORMATION: /product- "Module 5 of SorB"  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 46851..47891  
OTHER INFORMATION: /product- "Sorb"  
OTHER INFORMATION: /note- "The protein encoded by the sorB gene is highly  
OTHER INFORMATION: homologous to the methyltransferase from Streptomyces  
OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the  
OTHER INFORMATION: polyketide rapamycin."  
US-08-764-233A-1  
Query Match 3.9%; Score 47.6; DB 2; Length 49377;  
Best local similarity 44.1%; Pred No. 0.012; Mismatches 359; Indels 3; Gaps 2;  
Matches 286; Conservative 0;  
QY 148 GGCATGAGTCTCATCGAGCGAGTGGCGGAGAGGCTCGCTATCACCGACGCGAT 207  
Db 41609 GCCACGACGACTACCTCGCGGATGAGGCTATTCCTGCGCACACCGCGGCCAC 41668  
QY 208 TTGAAGCGCGAGCGCGCACTGTCGCGACCGCGACGAGTGTGGCGCAGCTGAT 267  
Db 41669 TTGAGCACCAGCGCGCTCTCGTAGCCACACCGGAGAGACTCTCTCCGCTCGAC 41728  
QY 268 TTATGCTCAAGGTCAAGACGATAGGCGGGAATAGCGCGCTGCGACGCGGACG 327  
Db 41729 TCCTGCGCGCGACCAAGCCCGCGCGGAGACGACGCTCTCGAGCGAGCGGAGCGACGCG 41788





APPLICANT: Crouzet, Joel  
 APPLICANT: Jacques, Nathalie  
 APPLICANT: Lacroix, Patricia  
 APPLICANT: Thibaut, Denis  
 APPLICANT: Zagorec, Monique  
 APPLICANT: Debussche, Laurent  
 APPLICANT: De Crecy-Lagard, Valerie  
 TITLE OF INVENTION: Polypeptides Involved in The  
 TITLE OF INVENTION: Biosynthesis of Streptogramins, Nucleotide Sequences  
 TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
 NUMBER OF SEQUENCES: 43  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
 STREET: 1300 I Street, N.W., Suite 700  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20005-3315  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/403,852D  
 FILING DATE: 10-MAY-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/FR 93/00923  
 FILING DATE: 25-SEP-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: FR 92/11441  
 FILING DATE: 25-SEP-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meyers, Kenneth J.  
 REGISTRATION NUMBER: 25,146  
 REFERENCE/DOCKET NUMBER: 03806.0054-00000  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 408-4000  
 TELEFAX: (202) 408-4400  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1833 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: CDNA  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: S.pristinaespiralis  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 103..1689  
 US-08-403-852D-6

Query Match 3.4%; Score 42.4; DB 3; Length 1833;  
 Best Local Similarity 50.5%; Pred. No. 0.086;  
 Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Db 533 TGCTGATGGGGGGGTCGCCGCTGCAACCGCCGACGTCGTGATCGCGCCGCA 592  
 413 TGATGCTCGGACTGTGTCTTCGGCGCGGGGTCCGCTGGCGGCTTCGCCAGGACT 472  
 593 CGCGCGGCTACACGACGACCGCCGACGACGCGATGGCGCGGCGGCGGCTTCGTTGAG 652  
 473 CGCGCCACTATCGCCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 532  
 653 ACATCAACATGACAAATTTGGCACTGACGCGGCGGAGTTTGGGCGCGGATCCACATC 712  
 533 CCACCTTCGCGTATATGAGGCTTCGACGACGACGACGACGACGACGACGACGACGAC 592  
 713 GCTACTCATCGCGCTACGAGCTCG 736

Db 593 TGTCGGAGCGCGCGACGCTACTCG 616

RESULT 15  
 US-08-173-508-3  
 Sequence 3, Application US/08173508  
 Patent No. 5616485  
 GENERAL INFORMATION:  
 APPLICANT: Battfeld, Daniel  
 APPLICANT: Butler, Michael J.  
 APPLICANT: Hadary, Dany  
 APPLICANT: Jenish, David  
 APPLICANT: Krieger, Timothy  
 TITLE OF INVENTION: STREPTOMYCES PROTEINASES AND IMPROVED  
 TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND  
 TITLE OF INVENTION: POLYPEPTIDES  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W.  
 CITY: Washington, D.C.  
 COUNTRY: USA  
 ZIP: 20007-5109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/173,508  
 FILING DATE: 23-DEC-1993  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BERT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 18740/125/CACO  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202 672 5300  
 TELEFAX: 202 672 5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2185 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 531..2069  
 FEATURE:  
 NAME/KEY: sig\_peptide  
 LOCATION: 531..902  
 FEATURE:  
 NAME/KEY: mat\_peptide  
 LOCATION: 903..2069  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 531..533  
 OTHER INFORMATION: /note="Met at position -124  
 OTHER INFORMATION: represents fmet"  
 US-08-173-508-3

Query Match 3.4%; Score 42.4; DB 1; Length 2185;  
 Best Local Similarity 47.4%; Pred. No. 0.091;  
 Matches 127; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

Db 525 CGCGGCTGTCTGATGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 584  
 1250 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1309  
 585 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 644

```

Db 1310 CCCCAGCGCTTCTCTACACCGAGGCCGTGCTGCGGCTCATGTGGCGGCGGAGCA 1369
QY 645 GGTCTAGACATCAACATCGACAAACTTCGGCAACTCGACCGAGTTCTGCGCGGAT 704
Db 1370 CGGCGTCGACGTGACCAACAACAGCTATTACACGACCGCGTGTACTTCAACTGCAAGGA 1429
QY 705 CCACACTCGGTACTCATCGGCGTACGAGCTCGAGGTGCGGTCAAAAGCTGCGACTGT 764
Db 1430 CGACCCCGACCGAGAGCGGCTCGTCGAGGCCGTCTCGGGGCTCCCGGTACGCGGAGAA 1489
QY 765 GATTGGGCGCGTCCCTGGTGCCAGGCGCC 792
Db 1490 GAAGGGCGCGGTCAACGTGCGCGGCGCC 1517

```

Search completed: June 22, 2000, 15:03:42  
 Job time: 17700 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:07:14 ; Search time 5541.94 Seconds

(without alignments)  
903.245 Million cell updates/sec

Title: US-09-362-485-3

Perfect score: 1235  
Sequence: 1 ATCTTCAGATTATATCAGAC.....ANGCAAGCATGATGTCGNC 1235

Scoring table: IDENTITY - NDC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: gb\_est1: \*  
21: gb\_est2: \*  
22: gb\_est3: \*  
23: gb\_est4: \*  
24: gb\_est5: \*  
25: gb\_est6: \*  
26: gb\_est7: \*  
27: gb\_est8: \*  
28: gb\_est9: \*  
29: gb\_est10: \*  
30: gb\_est11: \*  
31: gb\_est12: \*  
32: gb\_est13: \*  
33: gb\_est14: \*  
34: gb\_est15: \*  
35: gb\_est16: \*  
36: gb\_est17: \*  
37: gb\_est18: \*  
38: gb\_est19: \*  
39: gb\_est20: \*  
40: gb\_est21: \*  
41: gb\_est22: \*  
42: gb\_est23: \*  
43: gb\_est24: \*  
44: gb\_est25: \*

45: gb\_est26: \*  
46: gb\_est27: \*  
47: gb\_est28: \*  
48: gb\_est29: \*  
49: gb\_est30: \*  
50: gb\_est31: \*  
51: gb\_est32: \*  
52: em\_est20: \*  
53: em\_est21: \*  
54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*  
59: gb\_est33: \*  
60: gb\_est34: \*  
61: gb\_est35: \*  
62: gb\_est36: \*  
63: gb\_est37: \*  
64: gb\_est38: \*  
65: em\_est27: \*  
66: em\_est28: \*  
67: em\_est29: \*  
68: em\_est30: \*  
69: gb\_est39: \*  
70: gb\_est40: \*  
71: gb\_est41: \*  
72: gb\_est42: \*  
73: gb\_est43: \*  
74: gb\_est44: \*  
75: em\_est31: \*  
76: em\_est32: \*  
77: em\_est33: \*  
78: em\_est34: \*  
79: gb\_est45: \*  
80: gb\_est46: \*  
81: gb\_est47: \*  
82: gb\_est48: \*  
83: gb\_est49: \*  
84: gb\_est50: \*  
85: gb\_est51: \*  
86: em\_est34: \*  
87: em\_est35: \*  
88: em\_est36: \*  
89: em\_est37: \*  
90: gb\_est52: \*  
91: gb\_est53: \*  
92: gb\_est54: \*  
93: gb\_est55: \*  
94: gb\_est56: \*  
95: em\_est38: \*  
96: em\_est39: \*  
97: em\_est40: \*  
98: em\_est41: \*  
99: em\_est42: \*  
100: em\_est43: \*  
101: em\_est44: \*  
102: gb\_est57: \*  
103: gb\_est58: \*  
104: em\_est45: \*  
105: gb\_est59: \*  
106: gb\_est60: \*  
107: gb\_est61: \*  
108: gb\_est62: \*  
109: gb\_est63: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES

No.	Score	Match	Length	DB	ID	Description
C 1	57.4	4.6	925	82	CNS0091P	AL053013 Drosophila
C 2	55.2	4.5	925	82	CNS0091P	AL053013 Drosophila
C 3	51.8	4.2	535	74	AV390505	AL053013 Drosophila
C 4	50.2	4.1	844	82	CNS0052P	AL056552 Drosophila
C 5	49.9	4.0	910	82	CNS0060N	AL056529 Drosophila
C 6	47	3.8	846	82	CNS010RJ	AL059337 Drosophila
C 7	46.5	3.8	1101	83	CNS0181E	AL108764 Drosophila
C 8	46.4	3.8	935	82	CNS006XK	AL066051 Drosophila
C 9	46.2	3.7	645	82	CNS01113	AL060051 Drosophila
C 10	45.2	3.7	935	82	CNS006XK	AL066051 Drosophila
C 11	45.4	3.7	591	69	AW128786	AL066051 Drosophila
C 12	45.2	3.7	744	83	CNS0172K	AL128768 Drosophila
C 13	45.2	3.7	1101	83	CNS0175Y	AL108460 Drosophila
C 14	45	3.6	315	51	AT1712257	AT1712257 Drosophila
C 15	45	3.6	846	82	CNS010RJ	AT1712257 Drosophila
C 16	44.8	3.6	942	82	CNS0072Q	AL059337 Drosophila
C 17	44.4	3.6	843	82	CNS00C51	AL056562 Drosophila
C 18	43.8	3.5	427	51	AT1736678	AT1736678 Drosophila
C 19	43.8	3.5	512	74	AV396794	AT1736678 Drosophila
C 20	43.8	3.5	521	74	AV397116	AV396794 Drosophila
C 21	43.8	3.5	526	74	AV394827	AV397116 Drosophila
C 22	43.6	3.5	1101	83	CNS016UE	AV394827 Drosophila
C 23	43.2	3.5	439	91	AO864063	AL107216 Drosophila
C 24	43.2	3.5	798	82	CNS00UJA	AL055851 Drosophila
C 25	43	3.5	413	74	AV396947	AV396947 Drosophila
C 26	41.8	3.4	414	80	AW286084	AV396947 Drosophila
C 27	41.8	3.4	465	80	AW287007	AW286084 Drosophila
C 28	41.8	3.4	530	74	AV396353	AW287007 Drosophila
C 29	41.8	3.4	839	82	CNS004NB	AV396353 Drosophila
C 30	41.8	3.4	932	82	CNS0072Q	AL054280 Drosophila
C 31	41.6	3.4	506	74	AW180713	AL066742 Drosophila
C 32	41.6	3.4	870	82	CNS006XJ	AL066742 Drosophila
C 33	41.4	3.4	1101	82	CNS00U6Z	AL0564271 Drosophila
C 34	41.4	3.4	1201	83	CNS014BJ	AL0768875 Drosophila
C 35	41.2	3.3	639	60	AT812194	AL013945 Drosophila
C 36	41.2	3.3	774	105	AQ3273321	AT812194 Drosophila
C 37	41.2	3.3	910	82	CNS0060N	AQ3273321 Drosophila
C 38	41.1	3.3	605	46	AT461529	AL056529 Drosophila
C 39	41.1	3.3	1101	83	CNS0157Y	AL056529 Drosophila
C 40	40.8	3.3	384	64	AQ054773	AT461529 Drosophila
C 41	40.8	3.3	385	21	D48746	AM054773 Drosophila
C 42	40.8	3.3	437	74	AV395790	AM054773 Drosophila
C 43	40.6	3.3	506	74	AV396405	D48746 Drosophila
C 44	40.4	3.3	574	91	AO848186	AV396405 Drosophila
C 45	40.2	3.3	916	83	CNS015SO	AO848186 Drosophila

## ALIGNMENTS

RESULT	1
CNS0091P/c	
LOCUS	CNS0091P 925 bp DNA GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACH19916 of Rpci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL053013
VERSION	AL053013.1 GI:4934461
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Mecoptera; Neoptera; Endopterygota; Diptera; Brachycera; Plecoptera; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 925) Genoscope.
REFERENCE	Direct Submission
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage...:
JOURNAL	BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP), the BDGP is constructing a physical map of the *Drosophila* melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org>. The BDGP *Drosophila* melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoset in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPci-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

**FEATURES**  
**SOURCE**

BASE COUNT	120 a	61 c	61 g	172 t	511 others
ORIGIN					

Query Match	4.68;	Score 57.4;	DB 82;	Length 925;
Best Local Similarity	11.08;	Score 57.4;	DB 82;	Length 925;

Matches 43; Conservative 176; Mismatches 142; Indels 0; Gaps 0;

442 CCCCTGCTTGCCCCGATGAGCGAGTCGCCGTCGACTCGCCGCCAGGTTGGCGCTTAC 503

D6 920 SCSCSCSSBSCSSSMSTSSNSNBCSCCSSBBSSTTSMSSSSBBSSSCGSSSSSGTS 8611

502 CACCTG TCGCAACCAAGGGGCGGTGTCTGATGGCGGGGTGCCCGGCTCGAA 561

Db 860 SACVKN1SSCGCCGCMABCMSSSSSCCGSASARGVKVRASGGAGKRGSSGASA 801

562 CCGGGCCGACGTGCTGATGCGGGCCGACCGCGCGCTACAAAGCAGCCCGATCGCC 621

DB 800 SHSSSSACBSSSSSSSCASCSWASSSSSASRRRCGGCAGCGGASSRRSSSSSSSSASAGS 741

622 AACGGCAATGGGCGACCGTTACGGTTCTAGACATCAACATCGACAACCTTCGGCACTC 681

DD /40 VSSABSSSSSCSSVSCSSVASSMCCSBBSSSSASASSSSSSSSASCASCSCCCTWSC 681

882 GACGCCGAGTCTCGGCCCTTACACTCGCTACTCATCGGCCCTACGAGCTCGAGGGT 741

000 SCISASMSAKSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSGSACGBSMSSG 621

142 GCCGACACGAGCCGACCTGGTATTGGGGCCCTCTCGTGGCCAGGGCCGACAGGCC 801

020 98363V3A35GMBSSV5B5SGGNBSGS6GGCGVC6SGSSSSSGSGSSGGSGSVCSGSSGCMC 561

	C	N
--	D	C
	C	N

RESULT 2

LOCUS	CNS0091P	925 bp	DNA	GSS	03-JUN-1999
-------	----------	--------	-----	-----	-------------

BACR19D16 of RFC-98 library from *Drosophila melanogaster* (from

Accession AL053013

**KEYWORDS** GSS, country, social capital

ORGANISM *Drosophila melanogaster*

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

REFERENCE 1 (bases 1 to 925)







FEATURES	pBelOBAC11.	
SOURCE	1. 846 Location/Qualifiers	
	/organism="Drosophila melanogaster"	
	/plasmid="pBelOBAC11"	
	/db_xref="taxon:7227"	
	/clone_11b="DrosBAC"	
	/clone="BACN04N13"	
	/note="end : 77"	
BASE COUNT	300 a	47 c 40 g 175 t 284 others
ORIGIN		
Query Match	3.8%;	Score 47; DB 82; Length 846;
Best Local Similarity	14.0%;	Pred. 5.7;
Matches	30; Conservative	111; Mismatches 73; Indels 0; Gaps 0;
QY	995	CCGTCACCAACCGACAGATCCCTATAGCTCGACCTTCCGACATGCTGCGGGCGG 1054
DB	801	SSSSSTSSSSSTSTBTBSSTTTSSSTSSSTSSSTSSSTSSSTSS 742
OY	1055	CGNCGCGTGCAGATCCGGGACTAGCCAAAGGCTTTCGACGACGAGGGGCGTTACTCT 1114
DB	741	SASASTSTSTTTCGCBSSBSSTSGTSSSTSBRTBTBSSTSSBSBSSSSSTSTBTBS 682
OY	1115	CCGACGCGTGCGCAGCGACCTGGGGGCGCCCTTCACCGA TCCCGCAGCGTGCCTG 1174
DB	681	SSSSSTSTSSSTSTSTSSSSSBSTSTBSSTCTTSTTTSSSCSSSSSTSSSS 622
OY	1175	GACTCTCGGCGCGCTTCGTACGCGGACACACATC 1208
DB	621	SSSCBTSSCSCSSSSCCTCGCCSCCCCCCTC 588
RESULT	7	
LOCUS	CNS0181E/c	GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC BACN3/H05 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.	
ACCESSION	AL108764	
VERSION	AL108764.1 Gi:5629068	
KEYWORDS	GSS.	
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 1 to 1101) Genome.	
AUTHORS	Direct Submission	
TITLE	Submitted (23-JUL-1999) GenomeScope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genomescope.cns.fr	
JOURNAL	- Web : www.genomecope.cns.fr)	
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.abi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billand at CSHG (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.	
FEATURES	Location/Qualifiers	
source	1. 1101	
	/organism="Drosophila melanogaster"	
	/plasmid="pBelOBAC11"	
	/db_xref="taxon:7227"	
	/clone_11b="DrosBAC"	
	/clone="BACN3/H05"	
	/note="end : SP6"	
BASE COUNT	61 a 215 c 301 g 235 t 289 others	
ORIGIN		

Query Match	Similarity	3. %	Score	46.6	DB	83	Length	1101	
Best Local	Similarity	40.48	Pred.	No. 7.1	Indels	1	Gaps	1	
Matches	101	Conservative	40	Mismatches	109	Indels	1	Gaps	1
QY	389	CGGCGACACGTCATTTGCTTACGAGACCGTGTCCAGACCGCCGACGCGGCGACTACCCCTTGC	448						
Db	308	CGCCMCVVVVVGMMAASCCCCCAVASCSGCCSCSSCCSCSCSCCCSCBACSCSCSCG	249						
QY	449	TTCGCCCGCATGAGCGCAAGTCGCGCGTGCACATCGCCGCCACAGTTGGCTTACACCTTA	508						
Db	248	AAGMCSCMSGVGACACAGAGGGGAGACCCGSCCASCSSCCGSCSSCCSSSSS	189						
QY	509	TGCGAACCAGAGGGGCGCGCGTGTCTGTATGGCGGGGTGCGCGCGCTGACACCCGCGG	568						
Db	188	CGCGACCCCS :AGGGGCGCCSGCCGCCACCCCGCCSAGACVAGACCGCCAC	130						
QY	569	AGGTGSGTGTATTCGCGCGC :GACCCGCGCGGTACACGACGACCCCGCATCGCCACGGCA	628						
Db	129	AGCGMAAGGGGAGAGGGCGCGGAGACCCCGCCACCCCGCCCGCCACCGCGCGGCA	70						
QY	629	TGCGCGCGCAC	639						
Db	69	AGCGCCCAAC	59						
RESULT	8								
LOCUS	CNS006XK	935 bp	DNA	GSS	03-JUN-1999				
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14M09 of RPCT-98 library from drosophila melanogaster (fruit fly), genomic survey sequence.								
ACCESSION	AL066051								
VERSION	AL066051.1	GI:4945019							
KEYWORDS	GSS:								
SOURCE	fruit fly.								
ORGANISM	Drosophila melanogaster								
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.								
REFERENCE	1 (bases 1 to 935)								
TITLE	Genoscope.								
JOURNAL	Direct Submission								
COMMENT	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)								
	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .								
FEATURES	Location/Qualifiers								
source	1..935								
	/organism="Drosophila melanogaster"								
	/db_xref="taxon:7227"								
	/clone_lib="RPCT-98"								
	/clone="BACR14M09"								
	/note="end : 77"								
BASE COUNT	257 a 170 c 162 g 96 t 250 others								
ORIGIN									
Query Match	3.8%	Score	46.4	DB	82	Length	935		





```

FEATURES
  source      Location/Qualifiers
              1..744
              /organism="Drosophila melanogaster"
              /plasmid="pBelobAC11"
              /db_xref="taxon:7227"
              /clone_lib="DrosBAC"
              /clone="BACN37B10"
              /note="end : Sp6"

BASE COUNT   20 a   191 c   245 g   105 t   183 others
ORIGIN
Query Match      3.7%; Score 45.2; DB 83; Length 744;
Best Local Similarity 34.0%; Pred. No. 12;
Matches 87; Conservative 41; Mismatches 128; Indels 0; Gaps 0;

QY 329 TCTGTTCAGCTTCTTTCATTTGGCCGCTGACGCTGCTTGCACGATGCTTGTGATT 388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 191 YCTTGTGCGCTTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 389 CCGGCACACACTCAATTCCTACGACGCGCTCCAGACCGCGCGCGCGCGCGCGCG 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 251 YCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 449 TTGCCCCGATGACGAGTCCCGCGCTGACGCTGCGCGCGCGCGCGCGCGCGCGCG 508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 311 CGGCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 509 TCGGACCCAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 371 CCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 569 ACGTCGTGTCATCGG 584
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 431 NGGNGSGSGCTKTGG 446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
LOCUS      CNS0175Y      1101 bp      DNA      GSS      26-JUL-1999
DEFINITION BACN37108 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION  AL108460
VERSION     AL108460.1 GI:5628764
KEYWORDS   GSS.
SOURCE     fruit fly.
            Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1101)
            Genoscope.
            Direct Submission
            Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaut at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBelobAC11.

FEATURES
  source      Location/Qualifiers
              1..1101
              /organism="Drosophila melanogaster"
              /plasmid="pBelobAC11"
              /db_xref="taxon:7227"
              /clone_lib="DrosBAC"
              /clone="BACN37108"
              /note="end : Sp6"

```

```

BASE COUNT   254 a   176 c   160 g   152 t   359 others
ORIGIN
Query Match      3.7%; Score 45.2; DB 83; Length 1101;
Best Local Similarity 12.6%; Pred. No. 13;
Matches 40; Conservative 157; Mismatches 121; Indels 0; Gaps 0;

QY 420 CCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 755 CGASAGAGGSGHGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 814
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 480 CGCCCGCGAGGTTGGCGGCTTACACCTGATCGAACCCAGGGGCGCGCGCGCGCGCG 539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 815 CAAVASASASVMSKAVASAVASCAVASAGAVASSCCRSVASVSAASVSSVSS 874
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 540 GGGCGGGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 599
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 875 SSSVSAASVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSV 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 600 CTACACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 659
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 935 SVAAASVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSS 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 660 CATGACAACTTGGCGCACTGACCGCGAGTCTGGCGCGCGGATCCACACTGCTACTC 719
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 995 AAAAAAASVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSVSSV 1054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 720 ATCGGCTACGAGCTCGA 737
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1055 VSAVASASASVSSVA 1072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 14
LOCUS      AT172257      315 bp      mRNA      EST      02-FEB-2000
DEFINITION 605069D09.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
ACCESSION  AT172257
VERSION     AT172257.1 GI:5006195
KEYWORDS   EST.
SOURCE     Zea mays.
            Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Zea.
            1 (bases 1 to 315)
REFERENCE   1
            Walbot V.
            Maize ESTs from various cDNA libraries sequenced at Stanford
            University
            Unpublished (1999)
            On Jun 22, 1998 this sequence version replaced gi:3246668.
            Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 605069 row: D column: 09.

FEATURES
  source      Location/Qualifiers
              1..315
              /organism="Zea mays"
              /cultivar="Ohio43"
              /db_xref="taxon:4577"
              /clone_lib="605 - Endosperm cDNA library from Schmidt lab"
              /tissue_type="nucellar, embryo, and endosperm"
              /dev_stage="10-14 days post-pollination"
              /lab_host="DH5(alpha)"
              /note="Organ: kernel; Vector: pAD-GAL4-2'; Site:1: EcoRI;
              Site:2: XhoI; kernel endosperm cDNA library from Schmidt
              lab"
BASE COUNT   61 a   105 c   106 g   42 t   1 others

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:51:53 ; Search time 8627.09 Seconds

(Without alignments)  
-139,484 Million cell updates/sec

Title: US-09-362-485-4

Perfect score: 1237

Sequence: 1 ATCTTGACAGTTATCGAAC.....GGAGCGATGATCGCGC 1237

Scoring table: IDENTITY\_NDC

Gapop 10.0 , Gapept 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Genembi:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pi1:\*  
8: gb\_pi2:\*  
9: gb\_pi3:\*  
10: gb\_pi4:\*  
11: gb\_pi5:\*  
12: gb\_pi6:\*  
13: gb\_pi7:\*  
14: gb\_pi8:\*  
15: gb\_pi9:\*  
16: gb\_pi10:\*  
17: gb\_pi11:\*  
18: gb\_pi12:\*  
19: gb\_pi13:\*  
20: gb\_pi14:\*  
21: gb\_pi15:\*  
22: gb\_pi16:\*  
23: gb\_pi17:\*  
24: gb\_pi18:\*  
25: gb\_pi19:\*  
26: gb\_pi20:\*  
27: gb\_pi21:\*  
28: gb\_pi22:\*  
29: gb\_pi23:\*  
30: gb\_pi24:\*  
31: gb\_pi25:\*  
32: gb\_pi26:\*  
33: gb\_pi27:\*  
34: gb\_pi28:\*  
35: gb\_pi29:\*  
36: gb\_pi30:\*  
37: gb\_pi31:\*  
38: gb\_pi32:\*  
39: gb\_pi33:\*  
40: gb\_pi34:\*  
41: gb\_pi35:\*  
42: gb\_pi36:\*  
43: gb\_pi37:\*  
44: gb\_pi38:\*

45: gb\_pi39:\*  
46: gb\_pi40:\*  
47: gb\_pi41:\*  
48: gb\_pi42:\*  
49: gb\_pi43:\*  
50: gb\_pi44:\*  
51: gb\_pi45:\*  
52: gb\_pi46:\*  
53: gb\_pi47:\*  
54: gb\_pi48:\*  
55: gb\_pi49:\*  
56: gb\_pi50:\*  
57: gb\_pi51:\*  
58: gb\_pi52:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1237	100.0	1237	5 A87607	A87607 Sequence 4
2	1237	100.0	1237	5 A89747	A89747 Sequence 4
3	1237	100.0	56414	1 MTVO02	AL008967 Mycobacte
4	1234.4	99.8	1236	5 A87613	A87613 Sequence 10
5	1234.4	99.8	1236	5 A89753	A89753 Sequence 10
6	1230.4	99.5	1235	5 A87606	A87606 Sequence 3
7	1230.4	99.5	1235	5 A89746	A89746 Sequence 3
8	1228	99.3	1228	5 A87608	A87608 Sequence 5
9	1228	99.3	1228	5 A89748	A89748 Sequence 5
10	1221.4	98.7	1235	5 A89749	A89749 Sequence 6
11	1221.4	98.7	1235	5 A87611	A87611 Sequence 8
12	1220.8	98.7	1235	5 A89751	A89751 Sequence 8
13	1220.8	98.7	1235	5 A87605	A87605 Sequence 2
14	1219.4	98.6	1245	5 A89745	A89745 Sequence 2
15	1219.4	98.6	1245	5 A87604	A87604 Sequence 1
16	1219.4	98.6	1245	5 A89744	A89744 Sequence 1
17	1219.4	98.6	1245	5 A89744	A89744 Sequence 1
18	1219.4	98.6	1245	5 A89744	A89744 Sequence 1
19	1213.8	98.1	1229	5 A87610	A87610 Sequence 7
20	1213.8	98.1	1229	5 A89750	A89750 Sequence 7
21	1206.4	97.5	1209	5 A87612	A87612 Sequence 9
22	1199.4	97.0	1208	5 A89752	A89752 Sequence 9
23	1192.4	96.4	1194	2 MTU92472	U92472 Mycobacteri
24	698	56.4	720	5 A89759	A89759 Sequence 16
25	538.4	43.5	40745	1 SC151	AL109848 Streptomy
26	467.6	37.8	682	5 A89758	A89758 Sequence 15
27	467.6	37.8	682	5 A89762	A89762 Sequence 19
28	377	30.5	8046	2 AF049107	AF049107 Myxococcu
29	351.2	28.4	11514	2 AE002028	AE002028 Deinococc
30	331.4	26.8	2780	1 RLE238118	AJ238118 Rhizobium
31	330.8	26.7	1585	2 AF070716	AF070716 Vibrio pr
32	312.8	25.3	1736	1 BACALDH	M33399 B. steatoche
33	293	23.7	1477	2 AF070715	AF070715 Shewanell
34	288.4	23.3	2113	1 D37807	D37807 Phormidium
35	277.2	22.4	1125	5 E01847	E01847 DNA encodin
36	258	20.9	137740	1 D90900	D90900 Synecocyst
37	243.2	19.7	1776	1 AB013821	AB013821 Enterobac
38	243	19.6	1776	1 BACALDH	M33399 B. sphaericu
39	235.6	19.0	1709	1 BACALDH	L20916 Bacillus su
40	235.6	19.0	1709	1 BACALDH	L20916 Bacillus su
41	235.6	19.0	1709	1 BACALDH	L20916 Bacillus su
42	218.2	17.6	1714	2 AF070714	AF070714 Carnobact
43	107.2	8.7	10512	2 AE001565	AE001565 Helicobac
44	99.2	8.0	10512	2 AE000639	AE000639 Helicobac
45	74.2	6.0	3160	1 RRC05294	U05294 Rhodospirill

ALIGNMENTS

RESULT 1  
 LOCUS A87607 1237 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 4 from Patent WO9836089.  
 ACCESSION A87607  
 VERSION A87607.1 GI:6736247  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1237)  
 AUTHORS Flohe, L. and Singh, M.  
 TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source 1. 1237  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 236 a 394 c 386 g 221 t  
 ORIGIN

Query Match 100.0%; Score 1237; DB 5; Length 1237;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-170;  
 Matches 1237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTTGCAAGATTATTCGAACCTTTCTTCACTGAAAGCTACAGTATCGAGGGGTAATC 60  
 DB 1 ATCTTGCAAGATTATTCGAACCTTTCTTCACTGAAAGCTACAGTATCGAGGGGTAATC 60  
 QY 61 ATGGCGCTGCTATTCGACGAGACCAAAACAAGCAATTCGGGTGGCCATACCCCG 120  
 DB 61 ATGGCGCTGCTATTCGACGAGACCAAAACAAGCAATTCGGGTGGCCATACCCCG 120  
 QY 121 GCCGCGCTGCGGAACTAACCCGCTGCGCATGAGTGTCTATCCAGCAGGTGCCGA 180  
 DB 121 GCCGCGCTGCGGAACTAACCCGCTGCGCATGAGTGTCTATCCAGCAGGTGCCGA 180  
 QY 181 GAGGCGCTGGGTATACCGAGCGGATTCGAAGCGGCGGCAACTGTGGGCAAC 240  
 DB 181 GAGGCGCTGGGTATACCGAGCGGATTCGAAGCGGCGGCAACTGTGGGCAAC 240  
 QY 241 GCCGACCAAGTGTGGGCGCGACGCTGATTTATGCTCAAGTCAAGAAGCAAGATGCGG 300  
 DB 241 GCCGACCAAGTGTGGGCGCGACGCTGATTTATGCTCAAGTCAAGAAGCAAGATGCGG 300  
 QY 301 GAATACGGCGCCCTGCGGACAGGGGCAATCTTTGTCACGTTCTTGCAATTGGCGCGTCA 360  
 DB 301 GAATACGGCGCCCTGCGGACAGGGGCAATCTTTGTCACGTTCTTGCAATTGGCGCGTCA 360  
 QY 361 CGTCTTGACACGATGCGTGTGATTCGCGACCAAGTCAATTTGCTTCAGACCGTC 420  
 DB 361 CGTCTTGACACGATGCGTGTGATTCGCGACCAAGTCAATTTGCTTCAGACCGTC 420  
 QY 421 CAGACCGCGGACGCGGCACTACCCCTGCTTGCCTCCGATGAGCAAGTGCCTGCACTC 480  
 DB 421 CAGACCGCGGACGCGGCACTACCCCTGCTTGCCTCCGATGAGCAAGTGCCTGCACTC 480  
 QY 481 GCGCGCCAGGTGGCGCTTACCACTGATGAGAACCCCAAGGGGCGCGGCTGTCTATG 540  
 DB 481 GCGCGCCAGGTGGCGCTTACCACTGATGAGAACCCCAAGGGGCGCGGCTGTCTATG 540  
 QY 541 GCGCGGGTGGCGGCTGTGACCGCGGACGCTGTGTGATTCGCGCGCGGACACCGCGG 600  
 DB 541 GCGCGGGTGGCGGCTGTGACCGCGGACGCTGTGTGATTCGCGCGCGGACACCGCGG 600  
 QY 601 TACAAACGACCGCCGATCGCCAAAGGCGATGGGCGGACCGCTTACGTTTACGATCAAC 660  
 DB 601 TACAAACGACCGCCGATCGCCAAAGGCGATGGGCGGACCGCTTACGTTTACGATCAAC 660  
 QY 661 ATGCAACAACTTGGCAATCGACGCGAGTGTCTGGGCGGATCCACACTCCCTACTCA 720

DB 661 ATGCAACAACTTGGCAATCGACCGGAGTTCGCGGCGGATCCACACTCCCTACTCA 720  
 QY 721 TCGGCTTACAGCTCGAGGCTGCCGTCAAACGTGCGGACCTGTGTATGGGCGGCTCTG 780  
 DB 721 TCGGCTTACAGCTCGAGGCTGCCGTCAAACGTGCGGACCTGTGTATGGGCGGCTCTG 780  
 QY 781 GTGCGGCGGCGCAAGGACCAATTAAGTCTGCAATTCCTTCCGCGATAGAAACA 840  
 DB 781 GTGCGGCGGCGCAAGGACCAATTAAGTCTGCAATTCCTTCCGCGATAGAAACA 840  
 QY 841 GTGCGGCTAGTGTGATATAGCATGACAGGCGGCTGTTTCGAAGGCTCAGACCG 900  
 DB 841 GTGCGGCTAGTGTGATATAGCATGACAGGCGGCTGTTTCGAAGGCTCAGACCG 900  
 QY 901 ACACCTACAGCACCCCGAGCTTCGCGCTGACAGACAGCTGTTTACGTGCGGGAAC 960  
 DB 901 ACACCTACAGCACCCCGAGCTTCGCGCTGACAGACAGCTGTTTACGTGCGGGAAC 960  
 QY 961 ATGCGCGCTGCGGTGCGGAAGAGCTGACCTACGCGCTGACCAACGAGATGCCGTAT 1020  
 DB 961 ATGCGCGCTGCGGTGCGGAAGAGCTGACCTACGCGCTGACCAACGAGATGCCGTAT 1020  
 QY 1021 GTGCTGAGCTTGCAGACCATGCTGCGGCGGCTGCCGCTGCAATCCGCACTAGCC 1080  
 DB 1021 GTGCTGAGCTTGCAGACCATGCTGCGGCGGCTGCCGCTGCAATCCGCACTAGCC 1080  
 QY 1081 AAGGCTTTCAGACGACGAGAGGGCGTTACTGTCCGAAGGGGTGSCACGACCTGAGG 1140  
 DB 1081 AAGGCTTTCAGACGACGAGAGGGCGTTACTGTCCGAAGGGGTGSCACGACCTGAGG 1140  
 QY 1141 GTGCGCTTACCGACGCGCGGCGAGCTGCTGCGCTGACTCTGCGCGCTGTTACGCCGAG 1200  
 DB 1141 GTGCGCTTACCGACGCGCGGCGAGCTGCTGCGCTGACTCTGCGCGCTGTTACGCCGAG 1200  
 QY 1201 CACACGTCGAGGATTAAGGAGGAGCATGATGTGGCGCG 1237  
 DB 1201 CACACGTCGAGGATTAAGGAGGAGCATGATGTGGCGCG 1237

RESULT 2  
 LOCUS A89747 1237 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 4 from Patent WO9832862.  
 ACCESSION A89747  
 VERSION A89747.1 GI:6738281  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1237)  
 AUTHORS Flohe, L. and Singh, M.  
 TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
 JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source 1. 1237  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 236 a 394 c 386 g 221 t  
 ORIGIN

Query Match 100.0%; Score 1237; DB 5; Length 1237;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-170;  
 Matches 1237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTTGCAAGATTATTCGAACCTTTCTTCACTGAAAGCTACAGTATCGAGGGGTAATC 60  
 DB 1 ATCTTGCAAGATTATTCGAACCTTTCTTCACTGAAAGCTACAGTATCGAGGGGTAATC 60  
 QY 61 ATGGCGCTGCTATTCGACGAGACCAAAACAAGCAATTCGGGTGGCCATACCCCG 120

```

Db 61 ATGCGCGTTCGATTCGACGAGACCAAAAACAAGATTCCGGGTGCGCATCACCCG 120
Qy 121 GCCGCGTTCGCGGAATTAACCCGTCGTGGCCATGAGTGTCTATCTCAAGGAGTCCGGA 180
Db 121 GCCGCGTTCGCGGAATTAACCCGTCGTGGCCATGAGTGTCTATCTCAAGGAGTCCGGA 180
Qy 181 GAGGCTGCTATCACCGAGCGGATTTCAAGGCGGAGCGGCAACTGTGTGGCACC 240
Db 181 GAGGCTGCTATCACCGAGCGGATTTCAAGGCGGAGCGGCAACTGTGTGGCACC 240
Qy 241 GCCGACGAGTGTGGGCGGAGCTGATTTATTTGCTCAAGGTCAAGAACCATATACGCGC 300
Db 241 GCCGACGAGTGTGGGCGGAGCTGATTTATTTGCTCAAGGTCAAGAACCATATACGCGC 300
Qy 301 GAATACGCGCGCTTCGACGACGCGGAGATCTTTGTCACGTTCTTGCAATTTGGCCGCTCA 360
Db 301 GAATACGCGCGCTTCGACGACGCGGAGATCTTTGTCACGTTCTTGCAATTTGGCCGCTCA 360
Qy 361 CGTGTTCGACCGATGCTTTGATTCGCGACGCGGACGCGGATTTGCTTACGAGACCGTC 420
Db 361 CGTGTTCGACCGATGCTTTGATTCGCGACGCGGACGCGGATTTGCTTACGAGACCGTC 420
Qy 421 CAGACCGCGGAGCGGAGCTACCCCTGCTTCCCGGATGAGCGAAGTCCCGTCTGACTC 480
Db 421 CAGACCGCGGAGCGGAGCTACCCCTGCTTCCCGGATGAGCGAAGTCCCGTCTGACTC 480
Qy 481 GCCGCGCGAGTTCGCTTACCACTGATGGAACCCAAAGCGCGCGGTGTCTGCTATG 540
Db 481 GCCGCGCGAGTTCGCTTACCACTGATGGAACCCAAAGCGCGGTGTCTGCTATG 540
Qy 541 GCCGCGCGAGTTCGCTTACCACTGATGGAACCCAAAGCGCGGTGTCTGCTATG 540
Db 541 GCCGCGCGAGTTCGCTTACCACTGATGGAACCCAAAGCGCGGTGTCTGCTATG 540
Qy 601 TACAACGAGCGCGCATGCGCAAGCGGATGGCGGACGTTACGCTTACGATCAAC 660
Db 601 TACAACGAGCGCGCATGCGCAAGCGGATGGCGGACGTTACGCTTACGATCAAC 660
Qy 661 ATCAGCAAACTTCGCGCAATCGAGCGCGGATTCGCGCGGATTCGCACTGCTACTCA 720
Db 661 ATCAGCAAACTTCGCGCAATCGAGCGCGGATTCGCGCGGATTCGCACTGCTACTCA 720
Qy 721 TCGGCGTACGAGTTCGAGGCGTCCGTCGCAAGCGTCCGACGCTGATTTGGGCGCTGTG 780
Db 721 TCGGCGTACGAGTTCGAGGCGTCCGTCGCAAGCGTCCGACGCTGATTTGGGCGCTGTG 780
Qy 781 GTGCGAGCGCGCAAGCGCACCATAATCTGCAATCTGTCGCGCATATGAAACCA 840
Db 781 GTGCGAGCGCGCAAGCGCACCATAATCTGCAATCTGTCGCGCATATGAAACCA 840
Qy 841 GGTGCGTACGAGTTCGAGATATAGCATGACGAGGCGGCTGTCTGAAAGGCTACGAGCG 900
Db 841 GGTGCGTACGAGTTCGAGATATAGCATGACGAGGCGGCTGTCTGAAAGGCTACGAGCG 900
Qy 901 ACCACGACGACGACGCGGATTCGCGCTGACGACGAGCGCTGTTTACTGCGTGGCGAAC 960
Db 901 ACCACGACGACGACGCGGATTCGCGCTGACGACGAGCGCTGTTTACTGCGTGGCGAAC 960
Qy 961 ATGCGCGGCTTCGAGGCGGAGGAGCGTACGCGGCTGACGAGGAGCGAGCATCCGAT 1020
Db 961 ATGCGCGGCTTCGAGGCGGAGGAGCGTACGCGGCTGACGAGGAGCGAGCATCCGAT 1020
Qy 1021 GTGCTGAGCTTCGCGACATGCTGCGGCGGCGGCTGCGGATTCGCGGATTCGCGGATTC 1080
Db 1021 GTGCTGAGCTTCGCGACATGCTGCGGCGGCGGCTGCGGATTCGCGGATTCGCGGATTC 1080
Qy 1081 AAAGGTCTTTCGAGGAGGAGGCGGCTTACTGTCGGAAGCGGTGGCGACGACTGGGG 1140
Db 1081 AAAGGTCTTTCGAGGAGGAGGCGGCTTACTGTCGGAAGCGGTGGCGACGACTGGGG 1140
Qy 1141 GTGCGCTTCACGAGCGCGGAGGCTGCTGAGCTGAGCTGCGCGGCTGCTTACGCGGAG 1200
Db 1141 GTGCGCTTCACGAGCGCGGAGGCTGCTGAGCTGAGCTGCGCGGCTGCTTACGCGGAG 1200

```

```

Qy 1201 CACAGCTCGGAGTAAGGAAGCGATGATGCGCGC 1237
Db 1201 CACAGCTCGGAGTAAGGAAGCGATGATGCGCGC 1237

```

```

RESULT 3
LOCUS MT002 56414 bp DNA BCT 17-JUN-1998
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.
ACCESSION AL008967
VERSION AL008967.1
KEYWORDS GI:3251491
SOURCE Mycobacterium tuberculosis.
ORGANISM Mycobacterium tuberculosis
Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:
Actinomycetales: Corynebacteriales: Mycobacteriaceae;
Mycobacterium.
1 (bases 1 to 56414)

```

```

REFERENCE
AUTHORS Cole, S.T., Brosch, P., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Comor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Najandaram, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squires, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.

```

```

TITLE Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence

```

```

JOURNAL Nature 393 (6685), 537-544 (1998)

```

```

MEDLINE 98295987
REMARK Erratum: [Published erratum appears in Nature 1998 Nov
12;396(6707):1901]

```

```

REFERENCE 2 (bases 1 to 56414)

```

```

AUTHORS Parkhill, J.
TITLE Direct Submission

```

```

JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk

```

```

COMMENT on Jun 27, 1998 this sequence version replaced g1:2624256.
Notes:

```

Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.

(URL: [http://www.sanger.ac.uk/Projects/M\\_tuberculosis/](http://www.sanger.ac.uk/Projects/M_tuberculosis/)) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.

Gene prediction was based on a Hidden Markov Model of TB genes implemented in TRAPNICE (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.

CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

#### FEATURES

```

Location/Qualifiers
1..56414
/organism="Mycobacterium tuberculosis"

```

```

/source "strain="H37Rv"
/db_xref="taxon:1773"

```

```

misc_feature <1..56414
/note="Fragment designated v002. Does not represent a
physical clone"

```

```

source <1..233
/organism="Mycobacterium tuberculosis"

```

```

/strain="H37Rv"
/db_xref="taxon:1773"

```

```

gene /clone="1154"
complement(3..527)

```

CDS  
 /gene="recx"  
 complement(3. .527)  
 /gene="recx"  
 /note="Rv2736c, (MTV002.01c), len: 174 aa. recx, similar to eg. RECX\_PSEAE P37860 regulatory protein recx from Pseudomonas aeruginosa (153 aa), fasta scores: opt: 161 z-score: 257.2 E(): 3.0e-07, 30.7% identity in 137 aa overlap. Overlaps and extends CDS from overlapping cosmid MTCY154.16c"  
 /codon\_start=1  
 /transl\_table=11  
 /product="recx"  
 /protein\_id="CA15532.1"  
 /db\_xref="GI:2624257"  
 /db\_xref="SPTREMBL:O33280"  
 /translation="MTVSCPSTSEREQARALCTLLTARSRTAEALAGLAKRGY PEDIGNVLDRLAAGLVDDTDFAEQWGRANAAKSKALAEHLAKVDDVITTT VLGIDGAGRGRAEKLVRARLRREVLIIDGCTDEANVSRLVAMLRGCGTLACEV VIAIAERERRRR"  
 complement(493. .2865)  
 /gene="recA"  
 complement(493. .2865)  
 /note="Rv2737c, (MTV002.02c), len: 790 aa. recA, identical to RECA\_MYCTU P26345 recA protein (790 aa). Contains self-splicing protein element (intron) from 2294 to 969 (c) similar to Intron II from TR:E332317 (EMBL:Y13030) DNA-directed DNA polymerase (EC 2.7.7.7) from Thermococcus sp. (1829 aa), fasta scores: opt: 81 z-score: 235.2 E(): 6e-06, 24.6% identity in 183 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop), PS00321 recA signature, and PS00819 protein splicing signature. See Davis et al. (1992) Cell 71(2):201-210"  
 /codon\_start=1  
 /transl\_table=11  
 /product="recA"  
 /protein\_id="CA15533.1"  
 /db\_xref="GI:2624258"  
 /db\_xref="SWISS-PROT:P26345"  
 /translation="MTQTPDRKALFLAQAQIKSYKGSVMRLGDEAQPISVPTG SIILDVAGIGGIPRGRVIEIYPPSSGKTITVALHVAANAQAAGVAATPDELHADP DYAKKLGVDTSLLVSPDTEQALIEIADMLIRSGALDIVIDSAVALPDELHEGM GDSHVGIQARLMSQALRKMTGALNNGTIAIFINQLRDKIVGFGSPETTGKALKE YASVRMDVGRVETLKDGTINAVGNIRVKKVKNKCLAGETRIEPTVGTTHREDVYK RPIHYVAACDKGTILHARPYVSWFDGTGDIVIGLRAGGIVATPDPHVLIEYGNRA AGELRKGDYRAPRRPFGSDASAPADHARILGYLIGDROGNGGKPIPIFNINOR ALIDVPTRIATIGCAHPQGRSLAIAHRPGRNVALDQQAIGYGLAMEKITPN WFEPTPIADIVGNLGLFESDGVWSREOTGALRVGYTTTSEQLAHQIHLMLRFV GSVTRDYDTPQKPSIVNGRIQSKRQVEVRISGMDNTAFSAESVPMGPRGAALIQ AIPBATOGRRRGSOATYLAEMTDVLANLDERGVTAQDAAMIGVASDGPGRKMOV LGASRLRRDRVQALADLDKFLHMLAEELRYSVIREVLPTRARTPFLVELELHTL VAGGVVHNSPPEKQAEFDILYKGISREGSLIDMGVGOGLIRKSGAETPTEGEGLS OGGENARNFLYEADVADAELEKKIKRKLIGAVYTDPSNDGYLAPVPF"  
 complement(790. .2112)  
 /gene="recA"  
 /note="recA intein"  
 complement(790. .807)  
 /gene="recA"  
 /note="PS00881 Protein splicing signature"  
 complement(2197. .2223)  
 /gene="recA"  
 /note="PS00321 recA signature"  
 complement(2644. .2667)  
 /gene="recA"  
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
 complement(2873. .2877)  
 /note="possible RBS upstream of Rv2737c"  
 complement(3247. .3453)  
 /gene="Rv2738c"  
 complement(3247. .3453)  
 /gene="Rv2738c"  
 /note="Rv2738c, (MTV002.03c), len: 68 aa. Unknown, N-terminus is highly similar to the N-terminus of the

upstream OREFMTV002.07c (78.4% identity in 37 aa overlap); also similar to AL020958|SC4H\_5 Streptomyces coelicolor cosmid 4H8 (64aa) opt: 185 z-score: 283.5 E(): 2.9e-08; 35.7% identity in 63 aa overlap"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein Rv2738c"  
 /protein\_id="CA15534.1"  
 /db\_xref="GI:2624260"  
 /db\_xref="SPTREMBL:O33281"  
 /translation="MAGVRLTEFHERVALHFGAAYGSSVLLDHLVTFGDSAAQAI EDGVEPDVWVRALCADFVPHDRW"  
 complement(3464. .4630)  
 /gene="Rv2739c"  
 complement(3464. .4630)  
 /note="Rv2739c, (MTV002.04c), len: 388 aa. Probable transferase, similar to eg. TR:Q51560 (EMBL:L28170) Pseudomonas aeruginosa thiamosyl transferase (426 aa), fasta scores: opt: 178 z-score: 226.3 E(): 1.9e-05, 25.9% identity in 425 aa overlap. Equivalent to Mycobacterium leprae protein MLCB33.02c (392 aa); fasta scores opt: 294723|MTCB33.2 Mycobacterium leprae cosmid B33 opt: 2112 z-score: 2364.5 E(): 0; 80.9% identity in 388 aa overlap  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein Rv2739c"  
 /protein\_id="CA15535.1"  
 /db\_xref="GI:2624261"  
 /db\_xref="SPTREMBL:O33282"  
 /translation="MRAVAVAGPDGSHSPALALCQRFRAADPTLTFTGVEMLEAR AAGIDAVELDLAATRDLDAGARLHRAAQAVALNPRALPELVSVITACGG MAALGLIPWELNPHPLPLPSKGLPAGSGLAGTIGRLRDAITKALGRSMRAG LROBAVAREIGLPPARDGPIPLRLIATIPALEVERDMPAAVAVVGLPHEPDRVLA IPAGCGPVVAVAPFATGCTAGTIGTVALOSITPGETVSSRLVSSRANGLVPPM AVAGIGSOAELLTRADIVICGGGCHGMAKTILAAVPMVYVGGGDQNEIARVROGS AVLIPLTADALVAANVNEVLSPPREARRAASVACADPVRVCHDALAC"  
 4674. .5123  
 /gene="Rv2740"  
 4674. .5123  
 /note="Rv2740, (MTV002.05), len: 149 aa. unknown, similar to M. leprae hypothetical protein TR:Q49650 (EMBL:294723) MLCB33.03 (82235\_83\_140) (178 aa), fasta scores: opt: 498 z-score: 538.9 E(): 7.3e-23, 51.6% identity in 161 aa overlap"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein Rv2740"  
 /protein\_id="CA15536.1"  
 /db\_xref="GI:2624262"  
 /db\_xref="SPTREMBL:O33283"  
 /translation="MAELTSPETPETTEAIRAVEAFNLALQNEDEPDVAALGDDL VYENVGFSIRIGRGRTATLIRMOGRGFEVYKIRIGDGAVALTEETDALLIGPLAV QEWVGVEVDGDRITLMDRYEDVYDMFKGLRLQVALVYPSLKATL"  
 5355. .6932  
 /gene="PE\_PGRS"  
 5355. .6932  
 /gene="PE\_PGRS"  
 /note="Rv2741, (MTV002.06), len: 525 aa. Member of M. tuberculosis PE\_PGRS subfamily, similar to many eg.

Query Match 100.0% Score 1237 DB 1 Length 56414  
 Best Local Similarity 100.0% Pred. No. 9.9e-171  
 Matches 1237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTTGCAATTAATGCAACTTCTTCATAGCAGCGTACGATCGAGGGGTAATC 60  
 |||||||  
 DB 38201 ATCTTGCAATTAATGCAACTTCTTCATAGCAGCGTACGATCGAGGGGTAATC 38260  
 |||||||  
 QY 61 ATGCGCGTGGTATTCGACGACGACCAAAACGAATTCGGGTGGCATCACC 120  
 |||||||

Db 38261 ATGCGCGTGGATATTCGACGACGACCAAAACAAATTCGGGTGGGCATCACCCG 38320  
 QY 121 GCCGGGCTGGCGAATAACCCGCTGTCGCAATGAGTGTCTATCCAGGAGGTGCCGA 180  
 Db 38321 GCCGGGCTGGCGAATAACCCGCTGTCGCAATGAGTGTCTATCCAGGAGGTGCCGA 38380  
 QY 181 GAGGCGTGGCTATACGACGCGGATTCAGAGGCGGCGGCACTGTCGCGACC 240  
 Db 38381 GAGGCGTGGCTATACGACGCGGATTCAGAGGCGGCGGCACTGTCGCGACC 38440  
 QY 241 GCCGACGAGGTGTGGCGGACGCTGATTTATGCTCAAGGTCAAGAACCGATAGCGCGC 300  
 Db 38441 GCCGACGAGGTGTGGCGGACGCTGATTTATGCTCAAGGTCAAGAACCGATAGCGCGC 38500  
 QY 301 GAATAGCGCGGCTGGGACGCGGAGATCTTGTACGCTTCTGATTTGGCGGCTCA 360  
 Db 38501 GAATAGCGCGGCTGGGACGCGGAGATCTTGTACGCTTCTGATTTGGCGGCTCA 38560  
 QY 361 CGTGTTCGACGAGATGCTGTGTGATTCGCGACGACGCTCAATTCCTACGAGACCGTC 420  
 Db 38561 CGTGTTCGACGAGATGCTGTGTGATTCGCGACGACGCTCAATTCCTACGAGACCGTC 38620  
 QY 421 CAGACCGCGGACGCGGCACTACCCCTGCTTGGCCGATAGCGAATGCCCGGTGCACTC 480  
 Db 38621 CAGACCGCGGACGCGGCACTACCCCTGCTTGGCCGATAGCGAATGCCCGGTGCACTC 38680  
 QY 481 GCCGCCGAGTTGGCGCTTACCACTGATGCGAACCAGGAGGCGCGGCTGTGCTGATG 540  
 Db 38681 GCCGCCGAGTTGGCGCTTACCACTGATGCGAACCAGGAGGCGCGGCTGTGCTGATG 38740  
 QY 541 GCGGGGGTCCCGGCGCTGCAACCGCGGACGCTGTGTATGCGCGCGGACCGCGCGC 600  
 Db 38741 GCGGGGGTCCCGGCGCTGCAACCGCGGACGCTGTGTATGCGCGCGGACCGCGCGC 38800  
 QY 601 TCAACGCGCGCGGCTTGGCGAAGGCGGATGGCGGCGGATGAGTTTACGATCAAC 660  
 Db 38801 TCAACGCGCGCGGCTTGGCGAAGGCGGATGGCGGCGGATGAGTTTACGATCAAC 38860  
 QY 661 ATGCAAACTTTCGCACTGCAACGCGGAGTTCTGCGCGGATCCACACTGCGCTACTCA 720  
 Db 38861 ATGCAAACTTTCGCACTGCAACGCGGAGTTCTGCGCGGATCCACACTGCGCTACTCA 38920  
 QY 721 TCGGCTTACGAGCTCGAGGCTGCGCTCAAAAGTCCGCACTGCTATTTGGGCGCTCTG 780  
 Db 38921 TCGGCTTACGAGCTCGAGGCTGCGCTCAAAAGTCCGCACTGCTATTTGGGCGCTCTG 38980  
 QY 781 GTGCGAGGCGGCAAGGACCAATATGATCTGCAATTCCTGTCGCGATATGAACCA 840  
 Db 38981 GTGCGAGGCGGCAAGGACCAATATGATCTGCAATTCCTGTCGCGATATGAACCA 39040  
 QY 841 GGTGCGGCTACTGTGATATATGATATGACGACGAGGCGGCTGTTTGAAGGCTACAGCG 900  
 Db 39041 GGTGCGGCTACTGTGATATATGATATGACGACGAGGCGGCTGTTTGAAGGCTACAGCG 39100  
 QY 901 ACCACTTACGACACCGCGAGCTTGGCGGTCGACGACGCTGTTTACTGCGTGGGAGC 960  
 Db 39101 ACCACTTACGACACCGCGAGCTTGGCGGTCGACGACGCTGTTTACTGCGTGGGAGC 39160  
 QY 961 ATGCGCGCTCTGCTGCGAAGAGCTGACCTAGCGGCTGACCAAGCGAGATGCGGAT 1020  
 Db 39161 ATGCGCGCTCTGCTGCGAAGAGCTGACCTAGCGGCTGACCAAGCGAGATGCGGAT 39220  
 QY 1021 AAGAGTCTTTCACGACGAGAGGCGGCTTACTGTGCGAAGGCTGCGCACTGGGG 1140  
 Db 39221 AAGAGTCTTTCACGACGAGAGGCGGCTTACTGTGCGAAGGCTGCGCACTGGGG 39280  
 QY 1141 GTGCGGCTTACGAGCGCGGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
 Db 39281 GTGCGGCTTACGAGCGCGGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 39340  
 QY 1201 CACAGCTGGGAGTAAAGGAGCGATGATGTCGCGC 1237  
 Db 39341 GTGCGGCTTACGAGCGCGGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 39400

QY 1201 CACAGCTGGGAGTAAAGGAGCGATGATGTCGCGC 1237  
 Db 39401 CACAGCTGGGAGTAAAGGAGCGATGATGTCGCGC 39437  
 RESULT 4  
 A87613  
 LOCUS A87613 1236 bp DNA  
 DEFINITION Sequence 10 from Patent WO9836089.  
 ACCESION A87613  
 VERSION A87613.1 GI:6736253  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 unclassified.  
 unclassified.  
 unclassified.  
 REFERENCE  
 1 (bases 1 to 1236)  
 FLOHE, L. and Singh, M.  
 TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL  
 Patent: WO 9836089-A-20-AUG-1998  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 Location/Qualifiers  
 source  
 1..1236  
 /organism="unclassified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 236 a 395 c 385 g 220 t  
 ORIGIN  
 Query Match 99.8%; Score 1234.4; DB 5; Length 1236;  
 Best Local Similarity 99.9%; Pred. No. 5.4e-170;  
 Matches 1235; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ATCTGCGATTAATGCACTTTCTTCACTAGAACCTACAGTATCGAGGCGTATC 60  
 Db 1 ATCTTTCAGATTATGCACTTTCTTCACTAGAACCTACAGTATCGAGGCGTATC 60  
 QY 61 ATGCGJGTGCTATTCGACGAGACGCAAAACAGCAATTCGCGGTGGCATCACCCG 120  
 Db 61 ATGCGCGTGGTATTCGACGAGACGCAAAACAGCAATTCGCGGTGGCATCACCCG 120  
 QY 121 GCCGCGTCCGCGAATTAACCGCTGCTGCGCATGAGTGTCTATCAGAGGTCGCGA 180  
 Db 121 GCCGCGTCCGCGAATTAACCGCTGCTGCGCATGAGTGTCTATCAGAGGTCGCGA 180  
 QY 181 GAGGCGTGGCTATCAGCAGCGGATTTCAAGGCGGAGCGCGCAATCTGGGCAAC 240  
 Db 181 GAGGCGTGGCTATCAGCAGCGGATTTCAAGGCGGAGCGCGCAATCTGGGCAAC 240  
 QY 241 GCCGACGAGTGTGGGCGGACGCTGATTTATTCAGAGTCAAGAACCGATAGCGCG 300  
 Db 241 GCCGACGAGTGTGGGCGGACGCTGATTTATTCAGAGTCAAGAACCGATAGCGCG 300  
 QY 301 GAATAGCGCGGCTGCGACGCGGACGATCTTTGTTACGCTTCTTGATTTGGCGCTCA 360  
 Db 301 GAATAGCGCGGCTGCGACGCGGACGATCTTTGTTACGCTTCTTGATTTGGCGCTCA 360  
 QY 361 CGTGTTCGACGAGATGCTGTGATTCGCGGACGAGTAAATGCTTACGAGACCGTC 420  
 Db 361 CGTGTTCGACGAGATGCTGTGATTCGCGGACGAGTAAATGCTTACGAGACCGTC 420  
 QY 421 CAGACCGCGGACGCGGCACTACCCCTGCTTGGCCGATAGCGAATGCCCGGTGCACTC 480  
 Db 421 CAGACCGCGGACGCGGCACTACCCCTGCTTGGCCGATAGCGAATGCCCGGTGCACTC 480  
 QY 481 GCCGCCGAGTTGGCGCTTACCACTGATGCGAACCAGGAGGCGCGGCTGTGCTGATG 540  
 Db 481 GCCGCCGAGTTGGCGCTTACCACTGATGCGAACCAGGAGGCGCGGCTGTGCTGATG 540  
 QY 541 GCGGGGGTCCCGGCGCTGCAACCGGCGAGCTGCTGTGATTCGCGCGGACCGCGCGC 600  
 Db 541 GCGGGGGTCCCGGCGCTGCAACCGGCGAGCTGCTGTGATTCGCGCGGACCGCGCGC 600

```

OY 601 TACAGCAGCCCGCATGCGCAAGGCGGCGGACCGTTACGGTTCTAGACATCAAC 660
    |||||||
Db 601 TACAGCAGCCCGCATGCGCAAGGCGGCGGACCGTTACGGTTCTAGACATCAAC 660
OY 661 ATGACAAACTTGGGCAACTGACGCGGAGTTCTGCGGCGCGATCAGACTGCTACTCA 720
    |||||||
Db 661 ATGACAAACTTGGGCAACTGACGCGGAGTTCTGCGGCGCGATCAGACTGCTACTCA 720
OY 721 TCGGCTACGAGCTGAGAGGAGGCGGCTCAAAAGTCCGACCTGATGATGGGCGCTCTG 780
    |||||||
Db 721 TCGGCTACGAGCTGAGAGGAGGCGGCTCAAAAGTCCGACCTGATGATGGGCGCTCTG 780
OY 781 GTGCGAGCGCGCAAGGCAAGCCAAATTAGTCTGCAATTCGCTGCGCGATATGAAACA 840
    |||||||
Db 781 GTGCGAGCGCGCAAGGCAAGCCAAATTAGTCTGCAATTCGCTGCGCGATATGAAACA 840
OY 841 GGTGCGGACTGAGTATAGCATATGACCAAGGCGGCGCTGTTGCAAGGCTCACAGCCG 900
    |||||||
Db 841 GGTGCGGACTGAGTATAGCATATGACCAAGGCGGCGCTGTTGCAAGGCTCACAGCCG 900
OY 901 ACCACCTACGACACCGACGCTGCGCGCTGACAGACAGCTGTTTACTGCGTGGCGAAC 960
    |||||||
Db 901 ACCACCTACGACACCGACGCTGCGCGCTGACAGACAGCTGTTTACTGCGTGGCGAAC 960
OY 961 ATGCGCGCTGCGTGGCGCGCAAGAGCTGACCTTACGCGCTGACCAAGCGACATGCGCTAT 1020
    |||||||
Db 961 ATGCGCGCTGCGTGGCGCGCAAGAGCTGACCTTACGCGCTGACCAAGCGACATGCGCTAT 1020
OY 1021 GTGCTGAGAGCTGCGGACGATGGTGGGCGGCGGCGTCCGCTGCAATCCGCACTAGCC 1080
    |||||||
Db 1021 GTGCTGAGAGCTGCGGACGATGGTGGGCGGCGGCGTCCGCTGCAATCCGCACTAGCC 1080
OY 1081 AAAGGCTTTTGGACGACGACAGAGGCGGCTTACTGTCCGAACGGGTGCGCACTGCGGG 1140
    |||||||
Db 1081 AAAGGCTTTTGGACGACGACAGAGGCGGCTTACTGTCCGAACGGGTGCGCACTGCGGG 1140
OY 1141 GTGCGGCTTACGAGGCGCGCGCGAGCTGCTGAGCTGACTGCGGCGCTGTTACCGCAG 1200
    |||||||
Db 1141 GTGCGGCTTACGAGGCGCGCGCGAGCTGCTGAGCTGACTGCGGCGCTGTTACCGCAG 1200
OY 1201 CACAGCTGCGGAGTAAAGGAGCGATGATGTCGCC 1236
    |||||||
Db 1201 CACAGCTGCGGAGTAAAGGAGCGATGATGTCGCC 1236

RESULT 5
A89753 1236 bp DNA PAT 22-JAN-2000
LOCUS A89753
DEFINITION Sequence 10 from Patent WO9832862.
ACCESSION A89753
VERSION A89753.1 GI:6738287
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1236)
AUTHORS Flohe, L. and Singh, M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
    source
        1..1236
        /organism="unidentified"
        /db_xref="taxon:32644"
BASE COUNT 236 a 395 c 385 g 220 t
ORIGIN
Query Match 99.8%; Score 1234.4; DB 5; Length 1236;
Best Local Similarity 99.9%; Pred. No. 5, 4e-170;
Matches 1235; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 ATCTTGCAATTAATGCAACTTCTTCACTGACGCTAGACGATATGAGAGGGGTAAATC 60

```

```

Db 1 ATCTTGCAATTAATGCAACTTCTTCACTGACGCTAGACGATATGAGAGGGGTAAATC 60
    |||||||
OY 61 ATGCGGCTGCTATTCGACGCGGAGACCAAAACAAAGATTCGGGGTGGCAATCAACCCG 120
    |||||||
Db 61 ATGCGGCTGCTATTCGACGCGGAGACCAAAACAAAGATTCGGGGTGGCAATCAACCCG 120
OY 121 GCCGGGTCGCGCACTAACCCGCTGCGCATGAGGTGCTATCCAGGAGAGTGGCGGA 180
    |||||||
Db 121 GCCGGGTCGCGCACTAACCCGCTGCGCATGAGGTGCTATCCAGGAGAGTGGCGGA 180
OY 181 GAGGCTCTGCGCTATACGACGCGGATTTCAAGGGCGGAGCGCGCAATGCTGGCACC 240
    |||||||
Db 181 GAGGCTCTGCGCTATACGACGCGGATTTCAAGGGCGGAGCGCGCAATGCTGGCACC 240
OY 241 GCCGACAGGTGTGGGCGGCGGCTGATTTATTGCTCAAGGCTAAAGAACGATAGCGCG 300
    |||||||
Db 241 GCCGACAGGTGTGGGCGGCGGCTGATTTATTGCTCAAGGCTAAAGAACGATAGCGCG 300
OY 301 GAATACGCGCGCTGCGACGACGCGGAGATCTGTACGTTCTTGCAATTTGGCGGCTCA 360
    |||||||
Db 301 GAATACGCGCGCTGCGACGACGCGGAGATCTGTACGTTCTTGCAATTTGGCGGCTCA 360
OY 361 CGTGTGACGCGGATGCTGTTGATTCGCGGACGACGCTCAATGCTTACGAGACGCTC 420
    |||||||
Db 361 CGTGTGACGCGGATGCTGTTGATTCGCGGACGACGCTCAATGCTTACGAGACGCTC 420
OY 421 CAGACGCGGACGCGGACGCTACCTGCTTGCCTGATGAGCGAAGTCCGCGGTGACTC 480
    |||||||
Db 421 CAGACGCGGACGCGGACGCTACCTGCTTGCCTGATGAGCGAAGTCCGCGGTGACTC 480
OY 481 GCGCGCGAGTGGCGGCTTACACCTGATGCGAACCAGGCGGCGGCTGCTGATG 540
    |||||||
Db 481 GCGCGCGAGTGGCGGCTTACACCTGATGCGAACCAGGCGGCGGCTGCTGATG 540
OY 541 GCGCGGCTGCGGCGGCTTGAACGCGGCGGAGCTGTGATGCGGCGGCGGCGGCGGCG 600
    |||||||
Db 541 GCGCGGCTGCGGCGGCTTGAACGCGGCGGAGCTGTGATGCGGCGGCGGCGGCGGCG 600
OY 601 TACAGCAGCCCGCATGCGCAAGGCGGATGGGCGGCGGCTTACGACTCAAC 660
    |||||||
Db 601 TACAGCAGCCCGCATGCGCAAGGCGGATGGGCGGCGGCTTACGACTCAAC 660
OY 661 ATGACAAACTTGGGCAACTGACGCGGAGTTCTGCGGCGCGATCAGACTGCTACTCA 720
    |||||||
Db 661 ATGACAAACTTGGGCAACTGACGCGGAGTTCTGCGGCGCGATCAGACTGCTACTCA 720
OY 721 TCGGCTACGAGCTGAGAGGAGGCGGCTCAAAAGTCCGACCTGATGATGGGCGCTCTG 780
    |||||||
Db 721 TCGGCTACGAGCTGAGAGGAGGCGGCTCAAAAGTCCGACCTGATGATGGGCGCTCTG 780
OY 781 GTGCGAGCGCGCAAGGCAAGCCAAATTAGTCTGCAATTCGCTGCGCGATATGAAACA 840
    |||||||
Db 781 GTGCGAGCGCGCAAGGCAAGCCAAATTAGTCTGCAATTCGCTGCGCGATATGAAACA 840
OY 841 GGTGCGGACTGAGTATAGCATATGACCAAGGCGGCGCTGTTGCAAGGCTCACAGCCG 900
    |||||||
Db 841 GGTGCGGACTGAGTATAGCATATGACCAAGGCGGCGCTGTTGCAAGGCTCACAGCCG 900
OY 901 ACCACCTACGACACCGACGCTGCGCGCTGACAGACAGCTGTTTACTGCGTGGCGAAC 960
    |||||||
Db 901 ACCACCTACGACACCGACGCTGCGCGCTGACAGACAGCTGTTTACTGCGTGGCGAAC 960
OY 961 ATGCGCGCTGCGTGGCGCGCAAGAGCTGACCTTACGCGCTGACCAAGCGACATGCGCTAT 1020
    |||||||
Db 961 ATGCGCGCTGCGTGGCGCGCAAGAGCTGACCTTACGCGCTGACCAAGCGACATGCGCTAT 1020
OY 1021 GTGCTGAGAGCTGCGGACGATGGTGGGCGGCGGCTCCGCTGCAATCCGCACTAGCC 1080
    |||||||
Db 1021 GTGCTGAGAGCTGCGGACGATGGTGGGCGGCGGCTCCGCTGCAATCCGCACTAGCC 1080
OY 1081 AAAGGCTTTTGGACGACGACAGAGGCGGCTTACTGTCCGAACGGGTGCGCACTGCGGG 1140
    |||||||

```

Db 1081 AAGGCTCTTCAGCAGCAGAGAGGGGCTTACTCTCCAAAGGGGTGGCCACCACTGGGG 1140

QY 1141 GTGCGTTACAGAGCCCGCCAGCGTGTGGCTGACTCTCGGCGCTGTACCGCGAG 1200

Db 1141 GTGCGTTACAGAGCCCGCCAGCGTGTGGCTGACTCTCGGCGCTGTACCGCGAG 1200

QY 1201 CACACGTGGAGTAAGGAGGAGATGATCGGCC 1236

Db 1201 CACACGTGGAGTAAGGAGGAGATGATCGGCC 1236

RESULT 6

LOCUS A87606 1235 bp DNA PAT 22-JAN-2000

DEFINITION Sequence 3 from Patent WO9836089.

ACCESSION A87606

VERSION A87606.1 GI:6736246

KEYWORDS

SOURCE

ORGANISM

REFERENCE 1 (bases 1 to 1235)

AUTHORS Flohe, L., and Singh, M.

TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE

JOURNAL Patent: WO 9836089-A 20-AUG-1998;

FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)

FEATURES

source 1..1235

Location/Qualifiers

/organism="unidentified"

/db\_xref="taxon:32644"

BASE COUNT 236 a 394 c 382 g 220 t 3 others

ORIGIN

Query Match 99.5%; Score 1230.4; DB 5; Length 1235;

Best Local Similarity 99.7%; Pred. No. 2,1e-169;

Matches 1231; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANCTTCAGATTATGAACTTCTCATCTGACGCTACAGTATGAGAGGGTAATC 60

Db 1 ATTCTCAGATTATGAACTTCTCATCTGACGCTACAGTATGAGAGGGTAATC 60

QY 61 ATGCGCGTGGTATTCGACGACGACCAAAACAGAAATCCGGGTGGCATCACCCG 120

Db 61 ATGCGCGTGGTATTCGACGACGACCAAAACAGAAATCCGGGTGGCATCACCCG 120

QY 121 GCCGGGTGGGACCTACCCCTGCTGGCATGAGGTGCTCATCGAGCAGGTGCCGA 180

Db 121 GCCGGGTGGGACCTACCCCTGCTGGCATGAGGTGCTCATCGAGCAGGTGCCGA 180

QY 181 GAGGGGTGGCTATCACCAGCGGAGATTCAAGGGGCAAGGGCGGACGCTGGGAC 240

Db 181 GAGGGGTGGCTATCACCAGCGGAGATTCAAGGGGCAAGGGCGGACGCTGGGAC 240

QY 241 GCGGACGAGTGTGGCCGACGCTGATTTATGCTCAAGGTCAAAAGACGATAGCGG 300

Db 241 GCGGACGAGTGTGGCCGACGCTGATTTATGCTCAAGGTCAAAAGACGATAGCGG 300

QY 301 GATTAAGGGCGCTCGGACGACGAGATCTTTACAGTCTTGTGATTTGGCCGCTCA 360

Db 301 GATTAAGGGCGCTCGGACGACGAGATCTTTACAGTCTTGTGATTTGGCCGCTCA 360

QY 361 CGTGTTCACGATCGTGTGATTCGCGACACAGT AATTGCTACAGACCTGTC 420

Db 361 CGTGTTCACGATCGTGTGATTCGCGACACAGT AATTGCTACAGACCTGTC 420

QY 421 CAGACGCGGAGGCGGACCTACCTGCTTGGCCGACGACGAGTCCGCGTGCAGTC 480

Db 421 CAGACGCGGAGGCGGACCTACCTGCTTGGCCGACGACGAGTCCGCGTGCAGTC 480

QY 481 GCGGCGGAGTGGCGCTTACACGATGATGCGAACCAAGGGGCGCGGTGTGCTGATG 540

Db 481 GCGGCGGAGTGGCGCTTACACGATGATGCGAACCAAGGGGCGCGGTGTGCTGATG 540

QY 541 GCGGCGGAGTGGCGCTTACACGATGATGCGAACCAAGGGGCGCGGTGTGCTGATG 600

Db 541 GCGGCGGAGTGGCGCTTACACGATGATGCGAACCAAGGGGCGCGGTGTGCTGATG 600

QY 601 TACAAGCAGACCCCGCATCCCAAGGACATGGGCGGACGCTTACGGTTTACACATCAAC 660

Db 601 TACAAGCAGACCCCGCATCCCAAGGACATGGGCGGACGCTTACGGTTTACACATCAAC 660

QY 661 ATCGACAACTTCGGCACTGACGCGCGAGTTCTGCGGCGGATCCACACTCGTACTCA 720

Db 661 ATCGACAACTTCGGCACTGACGCGCGAGTTCTGCGGCGGATCCACACTCGTACTCA 720

QY 721 TCGGCTACGAGCTGAGGCTGCCGCAACGCTGCGGACGCTGATTTGGGCGCTCTG 780

Db 721 TCGGCTACGAGCTGAGGCTGCCGCAACGCTGCGGACGCTGATTTGGGCGCTCTG 780

QY 781 GTGCGAGGCGCCAGGACCAAAATTAGTCTGAAATTCTGCGGATGAAACCA 840

Db 781 GTGCGAGGCGCCAGGACCAAAATTAGTCTGAAATTCTGCGGATGAAACCA 840

QY 841 GGTGGGTACTGTGTGATATGACCATGACAGGCGGCTGTTGGAAGGCTACAGACG 900

Db 841 GGTGGGTACTGTGTGATATGACCATGACAGGCGGCTGTTGGAAGGCTACAGACG 900

QY 901 ACCACCTACGACACCCGACGCTTCCGCGGACGACGACCTGTTTACGCTGGCGAAC 960

Db 901 ACCACCTACGACACCCGACGCTTCCGCGGACGACGACCTGTTTACGCTGGCGAAC 960

QY 961 ATGCCCGCTCGGTGCGCAAGACGCTGACCTACGCGCTGACCAAGCGAGATGCGGTAT 1020

Db 961 ATGCCCGCTCGGTGCGCAAGACGCTGACCTACGCGCTGACCAAGCGAGATGCGGTAT 1020

QY 1021 GTGCTGAGCTTCCGACCATGCTGGCGGCGGCGGTGCGGATTCGGCACTAGCC 1080

Db 1021 GTGCTGAGCTTCCGACCATGCTGGCGGCGGCGGTGCGGATTCGGCACTAGCC 1080

QY 1081 AAGGCTTTCGACGACGACGAGGGGCTTACTGTCGGAAGGCTGGCCACGACTGGGG 1140

Db 1081 AAGGCTTTCGACGACGACGAGGGGCTTACTGTCGGAAGGCTGGCCACGACTGGGG 1140

QY 1141 GTGCGTTACCGAGCCGACGCTGCTGGCTACTCTCGGCGCTGTACGCGGAG 1200

Db 1141 GTGCGTTACCGAGCCGACGCTGCTGGCTACTCTCGGCGCTGTACGCGGAG 1200

QY 1201 CACACGTGGAGTAAGGAGGAGATGATGCGC 1235

Db 1201 CACACGTGGAGTAAGGAGGAGATGATGCGNC 1235

RESULT 7

LOCUS A89746 1235 bp DNA PAT 22-JAN-2000

DEFINITION Sequence 3 from Patent WO9832862.

ACCESSION A89746

VERSION A89746.1 GI:6738280

KEYWORDS

SOURCE

ORGANISM

REFERENCE 1 (bases 1 to 1235)

AUTHORS Flohe, L., and Singh, M.

TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM

JOURNAL Patent: WO 9832862-A 30-JUL-1998;

FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)

FEATURES

source 1..1235

Location/Qualifiers

/organism="unidentified"

/db\_xref="taxon:32644"

BASE COUNT 236 a 394 c 382 g 220 t 3 others

ORIGIN



Query Match 99.5%; Score 1230.4; DB 5; Length 1235;  
 Best Local Similarity 99.7%; Pred. No. 2.1e-169;  
 Matches 1231; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 ATCTTGCAATTAATGCACTTCTCTCATCTAGCAAGCCGTACAGATGAGAGGGGTAATC 60
   |||||||
Db 1 ATCTTGCAATTAATGCACTTCTCTCATCTAGCAAGCCGTACAGATGAGAGGGGTAATC 60
QY 61 ATGCGCGTGGTATTCAGCAGACCAAAACAAACGAATTCGGGTGGCCATCAACCCG 120
   |||||||
Db 61 ATGCGCGTGGTATTCAGCAGACCAAAACAAACGAATTCGGGTGGCCATCAACCCG 120
QY 121 GCGGCGTGGGGAATTAACCCGTGTCGCAATGAGTCTCATCAGCAGAGTCCCGA 180
   |||||||
Db 121 GCGGCGTGGGGAATTAACCCGTGTCGCAATGAGTCTCATCAGCAGAGTCCCGA 180
QY 181 GAGGCGTGGCTATACACGACGCGGATTTCAAGGCGGACGCGCACTGGTGGCACC 240
   |||||||
Db 181 GAGGCGTGGCTATACACGACGCGGATTTCAAGGCGGACGCGCACTGGTGGCACC 240
QY 241 GCGGACAGAGTGGGCGCAGCCTGATTTATTTCTCAAGATCAAGAACCGATAGCGCG 300
   |||||||
Db 241 GCGGACAGAGTGGGCGCAGCCTGATTTATTTCTCAAGATCAAGAACCGATAGCGCG 300
QY 301 GAATACGCGCGCTGGAGACAGGCGAGATCTTGTCAGCTCTTGATTTGGCCCGCTCA 360
   |||||||
Db 301 GAATACGCGCGCTGGAGACAGGCGAGATCTTGTCAGCTCTTGATTTGGCCCGCTCA 360
QY 361 CGTCTTGACAGGATGCGTTTGGATTCGGCAGCAGTCAATTTGCTACAGACCGTC 420
   |||||||
Db 361 CGTCTTGACAGGATGCGTTTGGATTCGGCAGCAGTCAATTTGCTACAGACCGTC 420
QY 421 CAGACCGCGGAGGCGGACATACCCCTGTTGCGCGATGAGGAAGTGCAGGTCGACTC 480
   |||||||
Db 421 CAGACCGCGGAGGCGGACATACCCCTGTTGCGCGATGAGGAAGTGCAGGTCGACTC 480
QY 481 GCGGCGCGAGTGGGCGCTTACCACTGATGCAACCCAAAGGCGCGGCTGCTGATG 540
   |||||||
Db 481 GCGGCGCGAGTGGGCGCTTACCACTGATGCAACCCAAAGGCGCGGCTGCTGATG 540
QY 541 GCGGCGGTCGCGGCGTGAACGCGGCGGCGCTGCTGATGAGGCGCGGCGCGCGG 600
   |||||||
Db 541 GCGGCGGTCGCGGCGTGAACGCGGCGGCGCTGCTGATGAGGCGCGGCGCGCGG 600
QY 601 TACAAACGAGCGCGCATGCGCAACGCGCATGGCGGCGGCGGCTTACGATCAAC 660
   |||||||
Db 601 TACAAACGAGCGCGCATGCGCAACGCGCATGGCGGCGGCGGCTTACGATCAAC 660
QY 661 ATGCAACAACTGCGCACTGAGCGCGGATGTCGCGCGGATCCACACTGCTACTCA 720
   |||||||
Db 661 ATGCAACAACTGCGCACTGAGCGCGGATGTCGCGCGGATCCACACTGCTACTCA 720
QY 721 TCGGCTACGAGCTGAGGCGTCCGTAACAGTGGCGACTGATTTGGGCGCTCTG 780
   |||||||
Db 721 TCGGCTACGAGCTGAGGCGTCCGTAACAGTGGCGACTGATTTGGGCGCTCTG 780
QY 781 GTCGCGAGCGCGCAACGCAAAATTAATGCTGCAATTTGCTGCGCATATGAACCA 840
   |||||||
Db 781 GTCGCGAGCGCGCAACGCAAAATTAATGCTGCAATTTGCTGCGCATATGAACCA 840
QY 841 GTCGCGTACTGTCGATATGATGCACTGACGAGGCGGCTGTTGGAAGGCTCAGACCG 900
   |||||||
Db 841 GTCGCGTACTGTCGATATGATGCACTGACGAGGCGGCTGTTGGAAGGCTCAGACCG 900
QY 901 ACCACTACGACACCGCAGCTTCCGTCAGACGACGCTTTTACTGGGTGGCGAAC 960
   |||||||
Db 901 ACCACTACGACACCGCAGCTTCCGTCAGACGACGCTTTTACTGGGTGGCGAAC 960
QY 961 ATGCGCGCTGGTGGCGCAAGGCTGCAAGCTGACGCGGCAAGGCGGCAAGGCGGAT 1020
   |||||||
Db 961 ATGCGCGCTGGTGGCGCAAGGCTGCAAGCTGACGCGGCAAGGCGGCAAGGCGGAT 1020
QY 1021 GTGTCGAGCTTGGCGGACCATGCTGCGGCGGCTGCGGTCGAATCCGGCACTACCC 1080

```

```

Db 1021 GTGTCGAGCTTGGCGGACCATGCTGCGGCGGCGGCTGCGGTCGAATCCGGCACTACCC 1080
QY 1081 AAAGCTCTTTGACCCAGAGAGGGCGTTACTGTCTCCAAAGGGTGGCCACGACCTGGGG 1140
   |||||||
Db 1081 AAAGCTCTTTGACCCAGAGAGGGCGTTACTGTCTCCAAAGGGTGGCCACGACCTGGGG 1140
QY 1141 GTGCGTTTACCGAGCGCGCCAGCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
   |||||||
Db 1141 GTGCGTTTACCGAGCGCGCCAGCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
QY 1201 CACACGTCGGGAGTAAGGAAGCATGATGTCGCC 1235
   |||||||
Db 1201 CACACGTCGGGAGTAAGGAAGCATGATGTCGNC 1235

RESULT 8
A87608
LOCUS A87608 1228 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 5 from Patent WO9836089.
ACCESSION A87608
VERSION A87608.1 GI:6736248
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.
REFERENCE
1 (bases 1 to 1228)
AUTHORS
Flohe, L. and Singh, M.
TITLE
TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL
Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
source
1..1228
Location/Qualifiers
BASE COUNT 236 a 391 c 382 g 219 t
ORIGIN

Query Match 99.3%; Score 1228; DB 5; Length 1228;
Best Local Similarity 100.0%; Pred. No. 4.6e-169;
Matches 1228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCTTGCAATTAATGCACTTCTCTCATCTAGCAAGCCGTACAGATGAGAGGGGTAATC 60
   |||||||
Db 1 ATCTTGCAATTAATGCACTTCTCTCATCTAGCAAGCCGTACAGATGAGAGGGGTAATC 60
QY 61 ATGCGCGTGGTATTCAGCAGACCAAAACAAACGAATTCGGGTGGCCATCAACCCG 120
   |||||||
Db 61 ATGCGCGTGGTATTCAGCAGACCAAAACAAACGAATTCGGGTGGCCATCAACCCG 120
QY 121 GCGGCGTGGGGAATTAACCCGTGTCGCAATGAGTCTCATCAGCAGAGTCCCGA 180
   |||||||
Db 121 GCGGCGTGGGGAATTAACCCGTGTCGCAATGAGTCTCATCAGCAGAGTCCCGA 180
QY 181 GAGGCGTGGCTATACGAGCGCGGATTTCAAGGCGGCGGCGGCAACTGGTGGCGACC 240
   |||||||
Db 181 GAGGCGTGGCTATACGAGCGCGGATTTCAAGGCGGCGGCGGCAACTGGTGGCGACC 240
QY 241 GCGGACAGAGTGGGCGCAGCCTGATTTATTTGCTCAAGGTCAAAGAACGATAGCGCG 300
   |||||||
Db 241 GCGGACAGAGTGGGCGCAGCCTGATTTATTTGCTCAAGGTCAAAGAACGATAGCGCG 300
QY 301 GAATACGCGCGCTGGAGACAGGCGAGATCTTGTCAGCTCTTGATTTGGCCCGCTCA 360
   |||||||
Db 301 GAATACGCGCGCTGGAGACAGGCGAGATCTTGTCAGCTCTTGATTTGGCCCGCTCA 360
QY 361 CGTCTTGACAGGATGCGTTTGGATTCGGCAGCAGTCAATTTGCTACAGACCGTC 420
   |||||||
Db 361 CGTCTTGACAGGATGCGTTTGGATTCGGCAGCAGTCAATTTGCTACAGACCGTC 420
QY 421 CAGACCGCGGAGGCGGACATACCCCTGTTGCGCGATGAGGAAGTGCAGGTCGACTC 480

```

Db 421 CAGACGGCCGACGGCGACCTACCCCTGCTGCCCCGATGAGCGAAGTGCAGGTC 480  
 QY 481 GCCGCCAGAGTTGGCGTTACCACTGATGCAACCAAGAGGGGGCGGTGTGATG 540  
 Db 481 GCCGCCAGAGTTGGCGTTACCACTGATGCAACCAAGAGGGGGCGGTGTGATG 540  
 QY 541 GCGCGGGTGGCGGGCGGTGCAACCGGCGCACTGTGTGATGCGCGCGCGCGCG 600  
 Db 541 GCGCGGGTGGCGGGCGGTGCAACCGGCGCACTGTGTGATGCGCGCGCGCGCG 600  
 QY 601 TACACGAGCGCGCGCATGCGCAAGGCGATGGGCGGCGGATGAGGTTCTAGATCAAC 660  
 Db 601 TACACGAGCGCGCGCATGCGCAAGGCGATGGGCGGCGGATGAGGTTCTAGATCAAC 660  
 QY 661 ATGCAAACTTGGGCAACTGCAACCGGAGTTCTGCGGCGCGATCAGACTGCTACTCA 720  
 Db 661 ATGCAAACTTGGGCAACTGCAACCGGAGTTCTGCGGCGCGATCAGACTGCTACTCA 720  
 QY 721 TCGGCTACGAGCTGAGGTTGGCGTCAAGTGGCGGACCTGCTGATGGGGCGCTCTG 780  
 Db 721 TCGGCTACGAGCTGAGGTTGGCGTCAAGTGGCGGACCTGCTGATGGGGCGCTCTG 780  
 QY 781 GTGCCAGCGCGCGCAAGGCGCAACCAATTAGTCTGCAATTGCTGCGCGCATATGAACA 840  
 Db 781 GTGCCAGCGCGCGCAAGGCGCAACCAATTAGTCTGCAATTGCTGCGCGCATATGAACA 840  
 QY 841 GGTGGGTACTGTGTGATATGACCATGCAACGAGGGCGCTGTTGGAAGGCTCAGACCG 900  
 Db 841 GGTGGGTACTGTGTGATATGACCATGCAACGAGGGCGCTGTTGGAAGGCTCAGACCG 900  
 QY 901 ACCACCTACGACCAACCGGAGTGGCGTGCAGACGAGTGTGTTTACTGGGTCGAGAC 960  
 Db 901 ACCACCTACGACCAACCGGAGTGGCGTGCAGACGAGTGTGTTTACTGGGTCGAGAC 960  
 QY 961 ATGCCCGCTGCTGGTCCGAAAGAGCTGACCTACGCGCTGACCAACGCGATGCGGTAT 1020  
 Db 961 ATGCCCGCTGCTGGTCCGAAAGAGCTGACCTACGCGCTGACCAACGCGATGCGGTAT 1020  
 QY 1021 GTGCTCGAGCTTGGCGACCATGCTGGGGGGGGGGGGTCCGGTTCGAATCCGGCACTACC 1080  
 Db 1021 GTGCTCGAGCTTGGCGACCATGCTGGGGGGGGGGGGTCCGGTTCGAATCCGGCACTACC 1080  
 QY 1081 AAAGGCTTTTGAGCAGCAAGAGGGGGCTTACTGTCGAAGGGGTGGCCAGCAGCTGGGG 1140  
 Db 1081 AAAGGCTTTTGAGCAGCAAGAGGGGGCTTACTGTCGAAGGGGTGGCCAGCAGCTGGGG 1140  
 QY 1141 GTGCCGTTACCGAGCGCGCGCAAGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
 Db 1141 GTGCCGTTACCGAGCGCGCGCAAGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
 QY 1201 CACACGTGGGAGTAAGGGAAGCGATGA 1228  
 Db 1201 CACACGTGGGAGTAAGGGAAGCGATGA 1228

## RESULT 9

A89748

Sequence 5 from Patent WO9832862.

PAT 22-JAN-2000

## LOCUS

A89748 1228 bp DNA

## DEFINITION

A89748

## VERSION

A89748.1

## KEYWORDS

unidentified.

## ORGANISM

unclassified.

## REFERENCE

1 (bases 1 to 1228)

## AUTHORS

Floehe, L. and Singh, M.

## TITLE

L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARIUM

## JOURNAL

Patent: WO 9832862-A 30-JUL-1998;

FLOEHE LEOPOLD (DE); SINGH MAHAVIR (DE)

Location/Qualifiers

1. 1228

/organism="unidentified"

BASE COUNT 236 a 391 c 382 g 219 t  
 ORIGIN  
 Query Match 99.3%; Score 1228; DB 5; Length 1228;  
 Best Local Similarity 100.0%; Pred. No. 4, 6e-169;  
 Matches 1228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTTGAGATTATCGACCTTCTTATCTAGAGCTACAGTATGAGAGGGTATC 60  
 Db 1 ATCTTGAGATTATCGACCTTCTTATCTAGAGCTACAGTATGAGAGGGTATC 60  
 QY 61 ATGCGCGTGGTATTCGACGAGCAACCAAAACCAAAATTCGGGTGCTATCACCCG 120  
 Db 61 ATGCGCGTGGTATTCGACGAGCAACCAAAACCAAAATTCGGGTGCTATCACCCG 120  
 QY 121 GCGGCGTGGGAGTATACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180  
 Db 121 GCGGCGTGGGAGTATACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180  
 QY 181 GAGGCTGCGCTATCACCGAGCGGATTTCAAGCGGCGAGCGCGCAACTGTGCGGACC 240  
 Db 181 GAGGCTGCGCTATCACCGAGCGGATTTCAAGCGGCGAGCGCGCAACTGTGCGGACC 240  
 QY 241 GCCGACGAGTGTGGCGGAGCTGATTTATGCTCAAGGTCAAAGAACGATAGCGGG 300  
 Db 241 GCCGACGAGTGTGGCGGAGCTGATTTATGCTCAAGGTCAAAGAACGATAGCGGG 300  
 QY 301 GAATAGCGCGCGCTGACACGCGGCAATCTGTTCACGTTTCTGCAATTGGCGCGTCA 360  
 Db 301 GAATAGCGCGCGCTGACACGCGGCAATCTGTTCACGTTTCTGCAATTGGCGCGTCA 360  
 QY 361 CGTCTTGACACCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420  
 Db 361 CGTCTTGACACCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420  
 QY 421 CAGACCGCGGAGCGGCGACCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
 Db 421 CAGACCGCGGAGCGGCGACCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
 QY 481 GCGCGCGAGGTTGGCGCTTACCACTGATGCAACCAAGGGGGCGGGTGTGATG 540  
 Db 481 GCGCGCGAGGTTGGCGCTTACCACTGATGCAACCAAGGGGGCGGGTGTGATG 540  
 QY 541 GCGCGGGTGGCGGGCGTGAACCGCGGAGCTGTGTGATCGGCGCGGACCGCGGCG 600  
 Db 541 GCGCGGGTGGCGGGCGTGAACCGCGGAGCTGTGTGATCGGCGCGGACCGCGGCG 600  
 QY 601 TACACGAGCGCGCATGCTGCAACCGGATGGCGCGGAGCTTACGCTTACGATCAAC 660  
 Db 601 TACACGAGCGCGCATGCTGCAACCGGATGGCGCGGAGCTTACGCTTACGATCAAC 660  
 QY 661 ATGCAAACTTGGCAACTGCAACCGGAGTTGCTGCAATTGCTGCGGCGATATGAACA 720  
 Db 661 ATGCAAACTTGGCAACTGCAACCGGAGTTGCTGCAATTGCTGCGGCGATATGAACA 720  
 QY 721 TCGGCTACGAGCTGAGAGGTGCGCTCAACGCTGCGCACTGTGTGATGGGGCGTCTG 780  
 Db 721 TCGGCTACGAGCTGAGAGGTGCGCTCAACGCTGCGCACTGTGTGATGGGGCGTCTG 780  
 QY 781 GTGCCAGGCGCGCAAGGCGCAACCAATTAGTCTGCAATTGCTGCGGCGATATGAACA 840  
 Db 781 GTGCCAGGCGCGCAAGGCGCAACCAATTAGTCTGCAATTGCTGCGGCGATATGAACA 840  
 QY 841 GTGCGGTACTGTGTGATATGACCATGCAACGAGCGGCTGTTGGAAGGCTCAGACCG 900  
 Db 841 GTGCGGTACTGTGTGATATGACCATGCAACGAGCGGCTGTTGGAAGGCTCAGACCG 900  
 QY 901 ACCACCTACGACCAACCGGAGTGGCGTGCAGACGAGCTGTTTACTGCTGCGGGAAC 960  
 Db 901 ACCACCTACGACCAACCGGAGTGGCGTGCAGACGAGCTGTTTACTGCTGCGGGAAC 960



JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
FEATURES Location/Qualifiers  
source 1..1235  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 235 a 395 c 384 g 220 t 1 others  
ORIGIN

Query Match 98.7% Score 1221.4; DB 5; Length 1235;  
Best Local Similarity 99.8%; Pred. No. 4.1e-168;  
Matches 1233; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```

QY 1 ATCTTGAGATTAAATGAACTTCTTCATCTAGAGCGTACAGATTCAGAGGGGTAATC 60
Db 1 ATCTTGAGATTAAATGAACTTCTTCATCTAGAGCGTACAGATTCAGAGGGGTAATC 60
QY 61 ATGCGCGTCCGATTCGCGAGACCAAAACAAAGAAATTCGGGTCCTACACCCG 120
Db 61 ATGCGCGTCCGATTCGCGAGACCAAAACAAAGAAATTCGGGTCCTACACCCG 120
QY 121 GCGGCGCTGCGGGAATTAACCCGCTGCGCATAGCTCATCAGGAGGTCCGGA 180
Db 121 GCGGCGCTGCGGGAATTAACCCGCTGCGCATAGCTCATCAGGAGGTCCGGA 180
QY 181 GAGGCGCTGCGCTATACCGAGCGGATTTCAAGCGCGCAACCTGGTGGCACC 240
Db 181 GAGGCGCTGCGCTATACCGAGCGGATTTCAAGCGCGCAACCTGGTGGCACC 240
QY 241 GCGGCGAGGTGGGCGGAGCGCTATTATTCGTCAGAGTCAAGAACCGG AGCGGCG 300
Db 241 GCGGCGAGGTGGGCGGAGCGCTATTATTCGTCAGAGTCAAGAACCGG AGCGGCG 300
QY 301 GAATACGCGCGCTCGACACAGCGGAGATCTTGTACGTTCTTTCATTTGGCCGCTCA 360
Db 301 GAATACGCGCGCTCGACACAGCGGAGATCTTGTACGTTCTTTCATTTGGCCGCTCA 360
QY 361 CGTCTTGCACCGATGCGCTGTTGATTCGCGACACAGTAAATGCTTACAGACCGTCC 420
Db 361 CGTCTTGCACCGATGCGCTGTTGATTCGCGACACAGTAAATGCTTACAGACCGTCC 420
QY 421 CAGACCGCGAGCGGCGCACTACCTGCTGGCCCGATGAGAGAAATGCGCGGTGACTC 480
Db 421 CAGACCGCGAGCGGCGCACTACCTGCTGGCCCGATGAGAGAAATGCGCGGTGACTC 480
QY 481 GCGGCGAGGTGGGCTTACCACTGATGCGAACCCAAAGGGGCGCGGTGCTGATG 540
Db 481 GCGGCGAGGTGGGCTTACCACTGATGCGAACCCAAAGGGGCGCGGTGCTGATG 540
QY 541 GCGGCGGTGCGCGGTGCGAACCGGCGGAGTCTGCTGATTCGCGCGGACCGCGCGC 600
Db 541 GCGGCGGTGCGCGGTGCGAACCGGCGGAGTCTGCTGATTCGCGCGGACCGCGCGC 600
QY 601 TACAACGAGCGCGGATCGCAACGCGATGGGCGGAGCGCTTACGCTTCTAGACATCAAC 660
Db 601 TACAACGAGCGCGGATCGCAACGCGATGGGCGGAGCGCTTACGCTTCTAGACATCAAC 660
QY 661 ATGCGCAAACTTGGCACTGAGAGCGGATTCGCGCGCGGATTCAGACATCGTACCA 720
Db 661 ATGCGCAAACTTGGCACTGAGAGCGGATTCGCGCGCGGATTCAGACATCGTACCA 720
QY 721 TCGGCGCTACGAGTCCGATGCGTAAAGCTCGGATTCGATGGGCGGCTGCTG 780
Db 721 TCGGCGCTACGAGTCCGATGCGTAAAGCTCGGATTCGATGGGCGGCTGCTG 780
QY 781 GTGCGAGCGCGCAAGGCGCAATTTAGTCTGAAATTCATCTGCGGCAATATAACCA 840
Db 781 GTGCGAGCGCGCAAGGCGCAATTTAGTCTGAAATTCATCTGCGGCAATATAACCA 840
QY 841 GGTGCGGCTACTGATGATATAGCATGAGCGGCGGCTTTTCGAAGGCTTCAGACCG 900
Db 841 GGTGCGGCTACTGATGATATAGCATGAGCGGCGGCTTTTCGAAGGCTTCAGACCG 900
QY 840 GGTGCGGCTACTGATGATATAGCATGAGCGGCGGCTTTTCGAAGGCTTCAGACCG 899

```

```

QY 901 ACCACTACGACGACCCGAGCTGCGCGTACAGACACACGCTGTTTACTGCTGGGCAAC 960
Db 901 ACCACTACGACGACCCGAGCTGCGCGTACAGACACACGCTGTTTACTGCTGGGCAAC 960
QY 961 ATGCGCGCTGCGGAGAGCTGACATCCGCTGACCAACGAGGAGATCCGCTAT 1020
Db 961 ATGCGCGCTGCGGAGAGCTGACATCCGCTGACCAACGAGGAGATCCGCTAT 1020
QY 1021 GTGCTCAGAGCTTCCGACATGCTGGCGGCGGCTGCGGCTGGAATCCGCGACTAGCC 1080
Db 1021 GTGCTCAGAGCTTCCGACATGCTGGCGGCGGCTGCGGCTGGAATCCGCGACTAGCC 1080
QY 1081 AAAGGCTTTCGAGCGACGAAAGGCGGCTTACTGTCCAGAGGCGGCGACCGACCTGGGG 1140
Db 1081 AAAGGCTTTCGAGCGACGAAAGGCGGCTTACTGTCCAGAGGCGGCGACCGACCTGGGG 1140
QY 1141 GTGCGCTTACCGAGCGCCCGGAGCGTGTGCTGCTGACTCTCGGCGCTGTTAGCGCGAG 1200
Db 1141 GTGCGCTTACCGAGCGCCCGGAGCGTGTGCTGCTGACTCTCGGCGCTGTTAGCGCGAG 1200
QY 1201 CACAGTCGCGGAGTAAAGGAGGATGATGCGGCGC 1236
Db 1201 CACAGTCGCGGAGTAAAGGAGGATGATGCGGCGC 1236

```

## RESULT 12

A87611 1235 bp DNA PAT 22-JAN-2000  
LOCUS Sequence 8 from Patent WO9836089.  
DEFINITION A87611.1 GI:6736251  
ACCESSION A87611.1  
VERSION A87611.1  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.

## REFERENCE

1 (bases 1 to 1235)  
AUTHORS Flohe/L. and Singh/M.  
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)

## FEATURES

source 1..1235  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 236 a 394 c 385 g 220 t  
ORIGIN

Query Match 98.7% Score 1220.8; DB 5; Length 1235;  
Best Local Similarity 99.8%; Pred. No. 5e-168;  
Matches 1233; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```

QY 1 ATCTTGAGATTAAATGAACTTCTTCATCTAGAGCGTACAGATTCAGAGGGGTAATC 60
Db 1 ATCTTGAGATTAAATGAACTTCTTCATCTAGAGCGTACAGATTCAGAGGGGTAATC 60
QY 61 ATGCGCGTCCGATTCGCGAGACCAAAACAAAGAAATTCGGGTCCTACACCCG 120
Db 61 ATGCGCGTCCGATTCGCGAGACCAAAACAAAGAAATTCGGGTCCTACACCCG 120
QY 121 GCGGCGGTGCGGGAATTAACCCGCTGCGCATAGCTCATCAGGAGGTCCGGA 180
Db 121 GCGGCGGTGCGGGAATTAACCCGCTGCGCATAGCTCATCAGGAGGTCCGGA 180
QY 181 GAGGCGTCCGCTATACCGAGCGGATTTCAAGCGGCGGAGCTGATGCTGCGCAC 240
Db 181 GAGGCGTCCGCTATACCGAGCGGATTTCAAGCGGCGGAGCTGATGCTGCGCAC 240
QY 241 GCGGACGAGTGTGGCGGCGGAGCTGATTTATGCTCAAGGTCAAAGACGATAGCGGCG 300
Db 241 GCGGACGAGTGTGGCGGCGGAGCTGATTTATGCTCAAGGTCAAAGACGATAGCGGCG 300

```

QY 301 GAATAGCGCCGCTGCGACAGCGGAGATCTGTTCAGCT..CTTGCAATTGGCCGGTGA 360  
 DB 301 GAATAGCGCCGCTGCGACAGCGGCG -GATCTGTTCAGCTTTTTCGATTTGGCCGGTGA 359  
 QY 361 GGTGCTTCAGCGAGATGCGTGTGGATTCCGCGACACCGTCAATTGCTTACGAGACCGTC 420  
 DB 360 GGTGCTTCAGCGAGATGCGTGTGGATTCCGCGACACCGTCAATTGCTTACGAGACCGTC 419  
 QY 421 CAGACCGCGAGCGGCGACTACCCCGTGTGCGCGATGAGGAGTGGCGCGTGGTGAAC 480  
 DB 420 CAGACCGCGAGCGGCGACTACCCCGTGTGCGCGATGAGGAGTGGCGCGTGGTGAAC 479  
 QY 481 GCGCGCCAGGTTGGCGCTTACCACTGATGCGAACCCAGGGGGCGCGGTGTGCTGATG 540  
 DB 480 GCGCGCCAGGTTGGCGCTTACCACTGATGCGAACCCAGGGGGCGCGGTGTGCTGATG 539  
 QY 541 GCGGGGGTCCCGGGGTGCAACCGCGCGCGCGTGTGATGATGGGGCGCGGACCGCGCG 600  
 DB 540 GCGGGGGTCCCGGGGTGCAACCGCGCGCGCGTGTGATGATGGGGCGCGGACCGCGCG 599  
 QY 601 TACAACGACGCCCGATCGCCCAACGCGATGGCGGCGACCGTTACGTTCTAGACATCAAC 660  
 DB 600 TACAACGACGCCCGATCGCCCAACGCGATGGCGGCGACCGTTACGTTCTAGACATCAAC 659  
 QY 661 ATCGCAAACTTCGCAACTCGACCGCGAGTGTGCGCGCGGATCCACACTGCTACTCA 720  
 DB 660 ATCGCAAA..CTTCGCAACTCGACCGCGAGTGTGCGCGCGGATCCACACTGCTACTCA 719  
 QY 721 TCGGCTTACGAGCTGAGAGGTGCGCTCAAAACGTGCGCACTGTTGATTGGGCGCTGCTG 780  
 DB 720 TCGGCTTACGAGCTGAGAGGTGCGCTCAAAACGTGCGCACTGTTGATTGGGCGCTGCTG 779  
 QY 781 GTGCCAGGCGCCAAAGGACCCCAATTAGTCTGGAATTCACTTGTGCGCATATGAACCA 840  
 DB 780 GTGCCAGGCGCCAAAGGACCCCAATTAGTCTGGAATTCACTTGTGCGCATATGAACCA 839  
 QY 841 GTGCGGTAAGTGTGATATAGCATATGACATGACGACGAGGCGGCT..TTTGAAGGGTCAAGACG 900  
 DB 840 GTGCGGTAAGTGTGATATAGCATATGACATGACGACGAGGCGGCTTTTGAAGGGTCAAGACG 899  
 QY 901 ACCACCTTACGACACCCGACGCTTCCGCTGACGACGACGCTGTTTTTACTGCGTGGCGAAC 960  
 DB 900 ACCACCTTACGACACCCGACGCTTCCGCTGACGACGACGCTGTTTTTACTGCGTGGCGAAC 959  
 QY 961 ATGCCCGCTGCTGCTGCGCAAGGCGTGAACCTGACGAGCGCTGACCAAGCGACGATGCCGAT 1020  
 DB 960 ATGCCCGCTGCTGCTGCGCAAGGCGTGAACCTGACGAGCGCTGACCAAGCGACGATGCCGAT 1019  
 QY 1021 GTGCTGAGACTTGGCGACATGCTGCGCGGCGCGCTGCCGTCGAAATCCGGCACTAGACC 1080  
 DB 1020 GTGCTGAGACTTGGCGACATGCTGCGCGGCGCGCTGCCGTCGAAATCCGGCACTAGACC 1079  
 QY 1081 AAAGTCTTTTGCAGCGACGAGAGGGCGTTACTGTCCGAAGGGGTGGCCACCGACTGGGG 1140  
 DB 1080 AAAGTCTTTTGCAGCGACGAGAGGGCGTTACTGTCCGAAGGGGTGGCCACCGACTGGGG 1139  
 QY 1141 GTGCGCTTACGAGCGCCGCGACGCTGCTGCGCTGACTCTGCGCGCGCTGTTAGCGCGAG 1200  
 DB 1140 GTGCGCTTACGAGCGCCGCGACGCGCTGCTGCGCTGACTCTGCGCGCGCTGTTAGCGCGAG 1199  
 QY 1201 CACACGTGCGGAGTAAAGGGAAGCGATGATGTCGGCC 1236  
 DB 1200 CACACGTGCGGAGTAAAGGGAAGCGATGATGTCGGCC 1235

RESULT 13  
 LOCUS AB9751 1235 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 8 from Patent W09832862.  
 ACCESSION AB9751  
 VERSION AB9751.1 GI:6738285  
 KEYWORDS  
 SOURCE unidentified.

ORGANISM unidentified  
 unclassified.  
 REFERENCE 1 (bases 1 to 1235)  
 AUTHORS Flohe, J. and Singh, M.  
 TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
 JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE LEOPOD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source 1..1235  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 236 a 394 c 385 g 220 t  
 ORIGIN

Query Match 98.7%; Score 1220.8; DB 5; Length 1235;  
 Best Local Similarity 99.8%; Pred. No. 5e-168;  
 Matches 1233; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 ATCTTGAGATTATCGACATTTCTCATCTGAGGCTACGATGAGAGGGTAAATC 60  
 DB 1 ATCTTGAGATTATCGACATTTCTCATCTGAGGCTACGATGAGAGGGTAAATC 60  
 QY 61 ATGCGCGTGGTATTCGACCGAGACCAAAACAGCAATTCCGGGTGGCCATCACCGG 120  
 DB 61 ATGCGCGTGGTATTCGACCGAGACCAAAACAGCAATTCCGGGTGGCCATCACCGG 120  
 QY 121 GCGGCGTCCGCGGAACTAACCCGCTGCGCATGAGGTGCTCATCCAGGAGGTGCCGA 180  
 DB 121 GCGGCGTCCGCGGAACTAACCCGCTGCGCATGAGGTGCTCATCCAGGAGGTGCCGA 180  
 QY 181 GAGGCTCGGCTATCACCGACGCGGATTTCAAGCGCGACGCGCGCACTGTCGCGAC 240  
 DB 181 GAGGCTCGGCTATCACCGACGCGGATTTCAAGCGCGACGCGCGCACTGTCGCGAC 240  
 QY 241 GCGGACAGGTGTGGCGCGACGCTGATTTATGCTCAAGGTCAAAGAACCGATAGCGCG 300  
 DB 241 GCGGACAGGTGTGGCGCGACGCTGATTTATGCTCAAGGTCAAAGAACCGATAGCGCG 300  
 QY 301 GAATAGCGCGCGCTGCGACAGCGGAGATCTGTGTCAGCTTGTGATTGGCGCGGTGA 360  
 DB 301 GAATAGCGCGCGCTGCGACAGCGGAGATCTGTGTCAGCTTGTGATTGGCGCGGTGA 359  
 QY 361 GTGCTTGCACCGATGCGTGTGGATTTCGCGACGACGTCATTTGCTTACGAGACCGTC 420  
 DB 360 GTGCTTGCACCGATGCGTGTGGATTTCGCGACGACGTCATTTGCTTACGAGACCGTC 419  
 QY 421 CAGACCGCGAGCGGCGACTACCCGCTGCTGCCCGATGAGCGCAATGCGCGGTGCACTC 480  
 DB 420 CAGACCGCGAGCGGCGACTACCCGCTGCTGCCCGATGAGCGCAATGCGCGGTGCACTC 479  
 QY 481 GCGCGCGAGTGTGGCGCTTACCACTGATGCGAACCCAGGGGGCGCGGTGTGCTGATG 540  
 DB 480 GCGCGCGAGTGTGGCGCTTACCACTGATGCGAACCCAGGGGGCGCGGTGTGCTGATG 539  
 QY 541 GCGGGGTGCGCGCGCGTGCAGACCGCGGACGCTGTGTGATGCGCGCGGACCGCGCG 600  
 DB 540 GCGGGGTGCGCGCGCGTGCAGACCGCGGACGCTGTGTGATGCGCGCGGACCGCGCG 599  
 QY 601 TACAACGACCGCGCATCGCCCAACGCGATGGCGGCGACCGTTACGTTCTAGACATCAAC 660  
 DB 600 TACAACGACCGCGCATCGCCCAACGCGATGGCGGCGACCGTTACGTTCTAGACATCAAC 659  
 QY 661 ATGACCAAACTTGGCACTGACGCGCGAATGTCGGCGCGGATCCACACTGCTACTCA 720  
 DB 660 ATGACCAAACTTGGCAAACTGACGCGCGAATGTCGGCGCGGATCCACACTGCTACTCA 719  
 QY 721 TCGGCTTACGAGCTGAGAGGTGCGCTCAAAACGTGCGCACTGTTGAGTGTGGGCGCTG 780  
 DB 720 TCGGCTTACGAGCTGAGAGGTGCGCTCAAAACGTGCGCACTGTTGAGTGTGGGCGCTG 779  
 QY 781 GTGCCAGCGCGCAAGGACCCCAATTAGTCTGGAATTCACTTGTGCGCATATGAACCA 840

Db 780 GTGCGAGGCGCCAGGACCAATTAAGTCTGCAATTCATTGCGGCATATGAAACA 839  
QY 841 GGTGCGGACTAGTGTGATATATAGCATGACAGGCGGCTGTTTCGAAGGCTCAGACCG 900  
Db 840 GGTGCGGACTAGTGTGATATATAGCATGACAGGCGGCTGTTTCGAAGGCTCAGACCG 899  
QY 901 ACCACCTACGACCAACCGGAGCTTGGCGGTGACACAGACGCTGTTTACTGCTGGCGAAC 960  
Db 900 ACCACCTACGACCAACCGGAGCTTGGCGGTGACACAGACGCTGTTTACTGCTGGCGAAC 959  
QY 961 ATGCGCGCTGCTGCGGAGAGACGCTGACGCTGACGCAACGAGCAGATGCGGTAT 1020  
Db 960 ATGCGCGCTGCTGCGGAGAGACGCTGACGCTGACGCAACGAGCAGATGCGGTAT 1019  
QY 1021 GTGCTGAGCTTGGCGACCATGCTGGCGGCGGCTGCCGCTGCAATCCGCACTACCC 1080  
Db 1020 GTGCTGAGCTTGGCGACCATGCTGGCGGCGGCTGCCGCTGCAATCCGCACTACCC 1079  
QY 1081 AAAGGCTTTTGAGAGCAGCAAGAGGCGGCTTACTGTCGAACGGGTGGCCACGACCTGGGG 1140  
Db 1080 AAAGGCTTTTGAGAGCAGCAAGAGGCGGCTTACTGTCGAACGGGTGGCCACGACCTGGGG 1139  
QY 1141 GTGCGGTTACACGAGCGCGGACGCTGCTGGCTGACTC\*GAGCGGCTGTTAGCGCGAG 1200  
Db 1140 GTGCGGTTACACGAGCGCGGACGCTGCTGGCTGACTC\*GAGCGGCTGTTAGCGCGAG 1199  
QY 1201 CACAGTCGGGAGTAAGGAGAGCGATGATGTCGGCC 1236  
Db 1200 CACAGTCGGGAGTAAGGAGAGCGATGATGTCGGCC 1235

RESULT 14  
A87605 1245 bp DNA PAT 22-JAN-2000  
LOCUS A87605 Sequence 2 from Patent WO9836089.  
DEFINITION A87605  
ACCESSION A87605.1 GI:6736245  
VERSION  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1245)  
AUTHORS Flohe, L. and Singh, M.  
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
FEATURES  
source 1..1245  
Location/Qualifiers  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 238 a 398 c 387 g 222 t  
ORIGIN

Query Match 98.6%; Score 1219.4; DB 5; Length 1245;  
Best Local Similarity 99.4%; Pred. No. 8e-168;  
Matches 1236; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

QY 235 GGCACCGCCGACAGGTGTGGCCGACGCTGATTATTTGCTCAAGGTCAAGAACGATA 294  
Db 241 GGCACCGCCGACAGGTGTGGCCGACGCTGATTATTTGCTCAAGGTCAAGAACGATA 300  
QY 295 GCGGCGGAATACGGCCCGCTGCGACACGGGAGATCTTTGTTACAGTTCTTGCAATTGGCC 354  
Db 301 GCGGCGGAATACGGCCCGCTGCGACACGGGAGATCTTTGTTACAGTTCTTGCAATTGGCC 360  
QY 355 GCGTACAGTGTGTCACAGATGCTTTTGG\*TTCCGGGACCAAGTCAATTTGCTTAGAG 414  
Db 361 GCGTACAGTGTGTCACAGATGCTTTTGGATTTCCGGGACCAAGTCAATTTGCTTAGAG 420  
QY 415 ACCGTCCAGACCGCCGACAGGCGGACACTACCCCTGCTTCCCGGATGAGAGGAGTCCCGGT 474  
Db 421 ACCGTCCAGACCGCCGACAGGCGGACACTACCCCTGCTTCCCGGATGAGAGGAGTCCCGGT 480  
QY 475 CGACTCGGCGCCAGGTTGGCGCTTACCACTGATGCGAACCAGAGGGGCGCGGTGTG 534  
Db 481 CGACTCGGCGCCAGGTTGGCGCTTACCACTGATGCGAACCAGAGGGGCGCGGTGTG 540  
QY 535 CTGATGGGCGGGGTGGCCGCGGCTGGAACCGGCGGACGTCGTTGATGCGGCGCGGACG 594  
Db 541 CTGATGGGCGGGGTGGCCGCGGCTGGAACCGGCGGACGTCGTTGATGCGGCGCGGACG 600  
QY 595 GCGGCTACAAACGACGCGCGCATCGCCACGATGGGCGGACCGTTACGTTCTAGAC 654  
Db 601 GCGGCTACAAACGACGCGCGCATCGCCACGATGGGCGGACCGTTACGTTCTAGAC 660  
QY 655 ATCAACATCGACAACTTGGGCACTGACGCGGAGTTCTGCGGCGCGGATCCACACTGC 714  
Db 661 ATCAACATCGACAACTTGGGCACTGACGCGGAGTTCTGCGGCGCGGATCCACACTGC 720  
QY 715 TACTCATGGGCTGAGAGCTGAGAGGTGGCTCAAGGTGGCGGACCTGTTGTTGGGGCC 774  
Db 721 TACTCATGGGCTGAGAGCTGAGAGGTGGCTCAAGGTGGCGGACCTGTTGTTGGGGCC 780  
QY 775 GTCTGTGTGCGGAGCGGCAAGGACCCCAATTACTCTGCAATTACTTGTGCGCATATG 834  
Db 781 GTCTGTGTGCGGAGCGGCAAGGACCCCAATTACTCTGCAATTACTTGTGCGCATATG 840  
QY 835 AAACAGGTGCGGTACTGCGATATAGCATGACAGGCGGCTGTTGGAAGGCTCA 894  
Db 841 AAACAGGTGCGGTACTGCGATATAGCATGACAGGCGGCTGTTGGAAGGCTCA 900  
QY 895 CGACCGACACCTACGACACCCGACGCTGCGCGGTGACACGACGCTTTTACTGGCTG 954  
Db 901 CGACCGACACCTACGACACCCGACGCTGCGCGGTGACACGACGCTTTTACTGGCTG 960  
QY 955 GCGAACATGCGCGGCTGCTGCGAAGACGTGACCTAGCGGCTGACCAACGCGACGATG 1014  
Db 961 GCGAACATGCGCGGCTGCTGCGAAGACGTGACCTAGCGGCTGACCAACGCGACGATG 1020  
QY 1015 CCGTATGTGCTGAGCTTGGCCGACATGCTGCGGCGGCGGCTGCCGATCCGCA 1074  
Db 1021 CCGTATGTGCTGAGCTTGGCCGACATGCTGCGGCGGCGGCTGCCGATCCGCA 1080  
QY 1075 GTAGCAAAAGTCTTTGAGACGCAAGAGGCTTACTGTTGCAAAAGGCTGAGGAG 1134  
Db 1081 GTAGCAAAAGTCTTTGAGACGCAAGAGGCTTACTGTTGCAAAAGGCTGAGGAG 1140  
QY 1135 CTGGGGGTGCGGCTTACCGAGACCGCCGACGCTGCTGAGCTTCTGCGCGCTGTTAC 1194  
Db 1141 CTGGGGGTGCGGCTTACCGAGACCGCCGACGCTGCTGAGCTTCTGCGCGCTGTTAC 1200  
QY 1195 GCGGACACAGCTGCGGAGTAAGGAGAGCATGATGTCGGCGG 1237  
Db 1201 GCGGACACAGCTGCGGAGTAAGGAGAGCATGATGTCGGCGG 1243

RESULT 15  
A89745 1245 bp DNA PAT 22-JAN-2000  
LOCUS A89745



DEFINITION Sequence 2 from Patent WO9832862.  
 ACCESSION A89745  
 VERSION A89745.1 GI:6738279  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1245)  
 FIOHE, L. and Singh, M.  
 L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
 TITLE Patent: WO 9832862-A 30-JUL-1998:  
 JOURNAL FIOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 Location/Qualifiers  
 1..1245  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 238 a 398 c 387 g 222 t  
 ORIGIN

Query Match 98.6% Score 1219.4; DB 5; Length 1245;  
 Best Local Similarity 99.4% Pred. No. 8e-168;  
 Matches 1236; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

```

QY 1 ATCTGAGATTAAATCGAACTTTCTTCACTAGAGGCTAGATGAGAGGGGTATC 60
DB 1 ATCTGAGATTAAATCGAACTTTCTTCACTAGAGGCTAGATGAGAGGGGTATC 60
QY 61 ATGGCGCTGCTATTCCTCCACGAGCAAAAACAAGTAATTCATTCGGGTGGCATC 120
DB 61 ATGGCGCTGCTATTCCTCCACGAGCAAAAACAAGTAATTCATTCGGGTGGCATC 120
QY 115 ACCCGGCGGCGTGGGAGGAACTAACCGCTGCGCCATGAGGCTCATCAGGCAAGT 174
DB 121 ACCCGGCGGCGTGGGAGGAACTAACCGCTGCGCCATGAGGCTCATCAGGCAAGT 180
QY 175 GCCGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCGCGCAATGCTC 234
DB 181 GCCGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCGCGCAATGCTC 240
QY 235 GGCACCGCGGCAAGGTTGGGCGGAGCTGATTTATTTGCTCAAGGTCMAAGAACCGATA 294
DB 241 GGCACCGCGGCAAGGTTGGGCGGAGCTGATTTATTTGCTCAAGGTCMAAGAACCGATA 300
QY 295 GCGCGGGAATAGCGCGCGCTCGACACGCGGAGATCTGTCACGTTCTTGCATTGGCC 354
DB 301 GCGCGGGAATAGCGCGCGCTCGACACGCGGAGATCTGTCACGTTCTTGCATTGGCC 360
QY 355 GCGTACGCTGCTTGCACCGATGCGTGTGGATTCGGGACACGATTCATTCCTACGAG 414
DB 361 GCGTACGCTGCTTGCACCGATGCGTGTGGATTCGGGACACGATTCATTCCTACGAG 420
QY 415 ACCTCCAGACCGCGCAAGGCGCACTACCGCTGCTGCGCGGATGAGCAAGTGGCGGT 474
DB 421 ACCTCCAGACCGCGCAAGGCGCACTACCGCTGCTGCGCGGATGAGCAAGTGGCGGT 480
QY 475 CGACTCGCGCGCGAGGTTGGGCTTACCACTGATGCGAACCAAGGGGGCGCGGTGTG 534
DB 481 CGACTCGCGCGCGAGGTTGGGCTTACCACTGATGCGAACCAAGGGGGCGCGGTGTG 540
QY 535 CTGATGGGCGGCGGCGCGGCTGGAACCGGCGCAAGTCTGCTGATCGCGCGCGCAC 594
DB 541 CTGATGGGCGGCGGCGGCTGGAACCGGCGCAAGTCTGCTGATCGCGCGCGCAC 600
QY 595 GCGCGCTACAAAGCGCGCGATCGCAACGCGATGGGCGGAGCGCTTACGCTTCTAGAC 654
DB 601 GCGCGCTACAAAGCGCGCGATCGCAACGCGATGGGCGGAGCGCTTACGCTTCTAGAC 660
QY 655 ATCAACATCGAAAGCTTGGGCAACTGACGCGGAGTTCTGGGCGGATTCACACTCGC 714
DB 661 ATCAACATCGAAAGCTTGGGCAACTGACGCGGAGTTCTGGGCGGATTCACACTCGC 720
QY 715 TACTCATCGGCTACGAGCTCGAGGTGCGGTCAAAAGTGGCGGACCTGGTATGGGGCC 774

```

```

DB 721 TACTCATCGGCTACGAGCTCGAGGCTGCCGCTAAACGTGCCGACTGCTGATGGGGCC 780
QY 775 GTCTGTGCTCCAGGCGCGCAAGSCACCCTAAATAGTCTGAATTCACCTTGTGCGCATATG 834
DB 781 GTCTGTGCTCCAGGCGCGCAAGSCACCCTAAATAGTCTGAATTCACCTTGTGCGCATATG 840
QY 835 AAACAGAGTGGCTAGTGGGATATAGCATGACAGGCGGCGGCTTTCGAAGGCTCA 894
DB 841 AAACAGAGTGGCTAGTGGGATATAGCATGACAGGCGGCGGCTTTCGAAGGCTCA 900
QY 895 CGACCGACACCTAGACACACCGGACGTTGCGGCTGACAGACAGGCTTTTACTGCGTG 954
DB 901 CGACCGACACCTAGACACACCGGACGTTGCGGCTGACAGACAGGCTTTTACTGCGTG 960
QY 955 GCGAACAATGCGCGCTCGGCTGCGGAGAGCTGAGCTACGCGCTGACCAAGCGAGATG 1014
DB 961 GCGAACAATGCGCGCTCGGCTGCGGAGAGCTGAGCTACGCGCTGACCAAGCGAGATG 1020
QY 1015 CCGTATGCTCGAGCTTGCACCATGCTGCGGCGGCGGCTGCGGCTGCAATCCGGCA 1074
DB 1021 CCGTATGCTCGAGCTTGCACCATGCTGCGGCGGCGGCTGCGGCTGCAATCCGGCA 1080
QY 1075 CTAGCCAAAGCTTTTCCGACGACGCAAGGCGGTTACTGTCCGAAGGCTGCGCACCGAC 1134
DB 1081 CTAGCCAAAGCTTTTCCGACGACGCAAGGCGGTTACTGTCCGAAGGCTGCGCACCGAC 1140
QY 1135 CTGGGGGTGCGCTTCCAGCGAGTCCGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 1194
DB 1141 CTGGGGGTGCGCTTCCAGCGAGTCCGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
QY 1195 GCGGACGACAGCTGCGGAGTAAGGGAAGCGATGATGCTGGCGG 1237
DB 1201 GCGGACGACAGCTGCGGAGTAAGGGAAGCGATGATGCTGGCGG 1243

```

Search completed: June 22, 2000, 14:52:44  
 Job time: 17579 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:06:27 ; Search time 458.59 seconds  
(without alignments)  
674.868 Million cell updates/sec

Title: US-09-362-485-4

Sequence: 1 ATCTTGCAGATTAAATCGAAC.....GGAAGCATGATGCGCCG 1237

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 6.3170

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N.Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1219.4	98.6	1245	1 V49626	Mycobacterium tube
2	1219.4	98.6	1260	1 V49510	Mycobacterium sp.
3	1219.4	98.6	1260	1 V49625	Mycobacterium tube
4	467.6	37.8	682	1 V49511	Mycobacterium mari
5	277.2	22.4	1125	1 N91423	Sequence of heat-r
6	212.4	17.2	28171	1 V52155	Streptococcus pneu
7	178.4	14.4	9280	1 V74442	Streptococcus aur
8	94.8	7.7	1074	1 T67971	H. pylori membrane
9	69.2	5.6	544	1 V43039	Streptococcus pneu
10	57.6	4.7	31096	1 V74370	Streptococcus pneu
11	50.8	4.1	390	1 O21833	Randomising oligon
12	50.8	4.1	390	1 O36859	PCR primer for 5'
13	50.4	4.1	1413	1 X34205	Mycobacterium spec
14	50.4	4.1	1722	1 X34206	Mycobacterium spec
15	50.4	4.1	15872	1 T68715	Streptomycetes venez
16	49	4.0	1269	1 Q03065	Sequence encoding
17	47.6	3.8	3346	1 T93610	Mycobacterium tube
18	47.6	3.8	28598	1 T06769	Sorangium cellulos
19	47.6	3.8	28958	1 T89956	The soraphen biosy
20	47.6	3.8	49377	1 V05287	Mycobacterium tube
21	46.4	3.8	985	1 V44439	Mycobacterium tube
22	46.4	3.8	985	1 V64548	M. tuberculosis im
23	46	3.7	1620	1 Q22482	griDL-1 gene codin
24	46	3.7	2682	1 T73117	Actinoplanes sp. a
25	46	3.7	2682	1 Q22485	griDL-1 gene. Reco
26	46	3.7	17955	1 V56642	Actinoplanes sp. a
27	45.4	3.7	5392	1 O64203	snb gene encoding
28	45.4	3.7	5392	1 O64201	Sequence compisin
29	42.4	3.4	1833	1 O64206	snb gene encoding
30	42.4	3.4	2185	1 V84066	Clone p5-4 encodin
31	42.4	3.4	2185	1 O9365	S. lividans protea
32	42.4	3.4	24379	1 T93095	Streptomycetes feno
33	42.4	3.4	24379	1 V25925	Streptomycetes roseo
34	42.2	3.4	329	1 V44425	Mycobacterium tube

35	42.2	3.4	329	1 V64534	M. tuberculosis im
36	42	3.4	3946	1 T93610	Mycobacterium tube
37	42	3.4	4018	1 O63879	Polyhydroxyalkanoa
38	41.8	3.4	882	1 V44403	Mycobacterium tube
39	41.8	3.4	882	1 V64512	M. tuberculosis im
40	41.6	3.4	12036	1 O04668	FHA structural gen
41	41.2	3.3	2414	1 O05926	Sequence encoding
42	41	3.3	29879	1 O46806	eryA region of S.
43	40.8	3.3	2151	1 O76252	HSV-2 protease, IC
44	40.8	3.3	2151	1 O76251	HSV-2 protease/ICP
45	40.8	3.3	2472	1 O84671	HSV-2 UL26 gene. N

## ALIGNMENTS

RESULT 1	
V49626	
ID V49626 standard; DNA; 1245 BP.	
AC V49626:	
DT 20-NOV-1998 (first entry)	
DE Mycobacterium tuberculosis	
KW ss: Alanine dehydrogenase: tuberculosis; TB; Mycobacteria; L-alanine.	
OS Mycobacterium tuberculosis.	
PN M09836089-82.	
PD 20-AUG-1998.	
PF 29-JAN-1998; E00483.	
PR 29-JAN-1997; EP-101338.	
PA (FLOH/) FLOHE L.	
PI Flohe L, Hutter B, Kolk A, Singh M;	
DR MPI: 98-457123/39.	
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity	
PT - useful for, e.g. for diagnosis, differentiation of strains,	
PT monitoring vaccination and identification of mycobacterial	
PT inhibitors	
PS Disclosure: Fig 3.19; 55pp; German.	
CC The Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in	
CC the production of kits for diagnosing tuberculosis (TB) and other for	
CC mycobacterial infections in humans or animals. Kits are used for direct	
CC diagnosis of TB on clinical samples (e.g. body fluids) and can	
CC differentiate between pathogenic and non-virulent strains, e.g. for	
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may	
CC also be used to identify substances that inhibit mycobacteria, for	
CC combating epidemics and for vaccination follow-up. Oligonucleotides	
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,	
CC also for culture confirmation of isolated strains and for chromosome	
CC fingerprinting to detect/differentiate between mycobacteria, and for	
CC L-alanine-specific biotransformation reactions. AlaDH is an early	
CC antigen, present extracellularly after only a few days of growth, making	
CC it an ideal drug target.	
SO Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;	

Query Match 98.6%; Score 1219.4; DB 1; Length 1245;  
Best Local Similarity 99.4%; Pred. No. 1.4e-245;  
Matches 1236; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

QY 1	ACTTGCAGATTAAATCGAAGCGGATTCGAGAGGGTATTC	60
DB 1	ACTTGCAGATTAAATCGAAGCGGATTCGAGAGGGTATTC	60
QY 61	ATGCGCGTGGTATTCGACGAGCAAAACACG-----AATTCGGGTGGCCATC	114
DB 61	ATGCGCGTGGTATTCGACGAGCAAAACACGAAATTCATTCGGGTGGCCATC	120
QY 115	ACCCGCGCGCGGCGGCGGAGTAACCCGTCGTCGATAGAGTCTCATCCAGCAGCT	174
DB 121	ACCCGCGCGCGGCGGCGGAGTAACCCGTCGTCGATAGAGTCTCATCCAGCAGCT	180
QY 175	GCGGAGAGGCGCTCGGATACGACGCGGATTCGAGGCGGCGGCGCACTGGTC	234
DB 181	GCGGAGAGGCGCTCGGATACGACGCGGATTCGAGGCGGCGGCGCACTGGTC	240

```

OY 235 GGCACCCGCGACAGGTGTGGGCGACGCTGATTATGTGCTCAAGGTCAAAGAACGATA 294
DB 241 GGCACCCGCGACAGGTGTGGGCGACGCTGATTATGTGCTCAAGGTCAAAGAACGATA 300
OY 295 GCGGCGGAATACGGCGCGCTGCGACACGGGCGAGATCTTTCACGTTCTTGCAATTTGGCC 354
DB 301 GCGGCGGAATACGGCGCGCTGCGACACGGGCGAGATCTTTCACGTTCTTGCAATTTGGCC 360
OY 355 GCGTCACGCTGTGACGCAATGCTGTGTGATTCGGGACCAAGGCGCATTTGGCTCGAG 414
DB 361 GCGTCACGCTGTGACGCAATGCTGTGTGATTCGGGACCAAGGCGCATTTGGCTCGAG 420
OY 415 ACCGTCAGACCGCGACGCGGACCTACCCCTGCTTCCCGCATGAGCGAAGTCGCGGT 474
DB 421 ACCGTCAGACCGCGACGCGGACCTACCCCTGCTTCCCGCATGAGCGAAGTCGCGGT 480
OY 475 CGACTCGCGCGCGAGTTGGGCTTACACGCTGATGCGAACCAGGGGCGCGGTG 534
DB 481 CGACTCGCGCGCGAGTTGGGCTTACACGCTGATGCGAACCAGGGGCGCGGTG 540
OY 535 CTGATGGGCGGCGCGCGCTGCGAACCAGGCGGACGCTGTGATGCGCGCGCGAC 594
DB 541 CTGATGGGCGGCGCGCGCTGCGAACCAGGCGGACGCTGTGATGCGCGCGCGAC 600
OY 595 GCGGCGTACACGACCGCGCATGCGCAACGGCATGGGCGCAACGCTTACGTTAGAC 654
DB 601 GCGGCGTACACGACCGCGCATGCGCAACGGCATGGGCGCGCAACGCTTACGTTAGAC 660
OY 655 ATCAATCATGACAAATCTGGGCACTGACGCGGAGTTCTGCGGCGGATTCACACTCGC 714
DB 661 ATCAATCATGACAAATCTGGGCACTGACGCGGAGTTCTGCGGCGGATTCACACTCGC 720
OY 715 TACTCATCGGCTTACGAGCTCGAGGTCGCTCAAAAGTGCAGACCTGTGATGGGCGC 774
DB 721 TACTCATCGGCTTACGAGCTCGAGGTCGCTCAAAAGTGCAGACCTGTGATGGGCGC 780
OY 775 GTCCTGTGCGAGGCGCGCAAGGCAATTAATGTCGAATTAATTCGCGGCGATG 834
DB 781 GTCCTGTGCGAGGCGCGCAAGGCAATTAATGTCGAATTAATTCGCGGCGATG 840
OY 835 AAACGAGTGGCTGATGCTGATATGATGACGACGAGGCGGCTGTTTCGAAGGCTCA 894
DB 841 AAACGAGTGGCTGATGCTGATATGATGACGACGAGGCGGCTGTTTCGAAGGCTCA 900
OY 895 CGACCGACCACTACGACCAACCGGAGTTCGCGCTGACAGACGCTGTTTACTGCTG 954
DB 901 CGACCGACCACTACGACCAACCGGAGTTCGCGCTGACAGACGCTGTTTACTGCTG 960
OY 955 GCGAATCATGCGCGCTGCGTGCAGAGAGCTGACCTACGCGCTGACCAACGCGAGATG 1014
DB 961 GCGAATCATGCGCGCTGCGTGCAGAGAGCTGACCTACGCGCTGACCAACGCGAGATG 1020
OY 1015 CCGTATGTGCTGAGCTTGGCGACATGCTGCGGCGGCGCTGCGGTCGAATCCGCA 1074
DB 1021 CCGTATGTGCTGAGCTTGGCGACATGCTGCGGCGGCGCTGCGGTCGAATCCGCA 1080
OY 1075 CTAGCCAAAGGCTTTCGACGACGAAAGGGGCTTACTGTCGAAAGGGTGGCGACCGCA 1134
DB 1081 CTAGCCAAAGGCTTTCGACGACGAAAGGGGCTTACTGTCGAAAGGGTGGCGACCGCA 1140
OY 1135 CTGGGGGTCGCTTACGACGAGCCGCGACGCTGCTGCTGACTGCGCGCTGTTAC 1194
DB 1141 CTGGGGGTCGCTTACGACGAGCCGCGACGCTGCTGCTGACTGCGCGCTGTTAC 1200
OY 1195 GCGGACACACGTCGGGATTAAGGGAAGGATGATTCGCGCG 1237
DB 1201 GCGGACACACGTCGGGATTAAGGGAAGGATGATTCGCGCG 1243

```

RESULT 2  
ID V49510 standard: DNA, 1260 BP.  
AC V49510;

```

DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlADH DNA.
KW Alanine dehydrogenase; AlADH; ADH; diagnosis; tuberculosis; pathogen;
OS Swimmers' disease; vaccine; epidemic; infection; identification; ss.
PN Mycobacterium sp.
PD MO9832862-A2.
PF 30-JUL-1998.
PR 29-JAN-1998; E00484.
PA 29-JAN-1997; EP-101339.
PI (FIOH/) FIOHE L.
PI Flohe L, Butler B, Kolk A, Singh M;
DR WPI: 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure: Page 11: 57pp. German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 98.6%; Score 1219.4; DB 1; Length 1260;
Best Local Similarity 99.4%; Pred. No. 1.4e-245;
Matches 1236; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

OY 1 ATCTTGACAGTAAATGCACTTCTCTCACTAGGAGTACGATGAGAGGGGTATC 60
DB 16 ATCTTGACAGTAAATGCACTTCTCTCACTAGGAGTACGATGAGAGGGGTATC 75
OY 61 ATGGCGCTGCTATTCGACCGAGACCAAAACACG-----AATTCGGGTGGCCATC 114
DB 76 ATGGCGCTGCTATTCGACCGAGACCAAAACACGAAATTCGAATTCGGGTGGCCATC 135
OY 115 ACCCGCGCGCGGCTGCGGAGTAAACCGTGTGGCATGAGTGTGTCATCCAGGCAAGT 174
DB 136 ACCCGCGCGCGGCTGCGGAGTAAACCGTGTGGCATGAGTGTGTCATCCAGGCAAGT 195
OY 175 GCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCGGCAACTGCTC 234
DB 196 GCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCGGCAACTGCTC 255
OY 235 GGCACCGCGGACAGGCTGCGGCTGAGCTATTTATGCTCAAGGTCGAAGAACGATA 294
DB 256 GGCACCGCGGACAGGCTGCGGCTGAGCTATTTATGCTCAAGGTCGAAGAACGATA 315
OY 295 GCGGCGGAATACGGCGCGCTGCGACAGGCGAGATCTTGTGCTCAAGGTCGAAGAACGATA 354
DB 316 GCGGCGGAATACGGCGCGCTGCGACAGGCGAGATCTTGTGCTCAAGGTCGAAGAACGATA 375
OY 355 GCGTACGCTGTTGACACGAGTCTGTTGATTCGCGACACGAGTCAATTTGCTACGAG 414
DB 376 GCGTACGCTGTTGACACGAGTCTGTTGATTCGCGACACGAGTCAATTTGCTACGAG 435
OY 415 ACCGTCAGACCGCGGACGCGGCGACTACCCCTGCTTCCCGCATGAGCGAAGTCGCGGT 474
DB 436 ACCGTCAGACCGCGGACGCGGCGACTACCCCTGCTTCCCGCATGAGCGAAGTCGCGGT 495
OY 475 CGACTCGCGCGCGAGTTGGGCTTACACCTGATGGAACCCAAAGGGGCGCGGTGTG 534
DB 496 CGACTCGCGCGCGAGTTGGGCTTACACCTGATGGAACCCAAAGGGGCGCGGTGTG 555
OY 535 CTGATGGGCGGCGGCTGCGGCTGGAACCGGCGAGCTGTGATGCGGCGCGGAC 594
DB 556 CTGATGGGCGGCGGCTGCGGCTGGAACCGGCGAGCTGTGATGCGGCGCGGAC 615

```

QY 595 GCGGCTACAAAGGAGCCGATCGCCAAAGGATGGGCGGACGCTTACGGTTCTAGAC 654  
 DB 616 GCGGCTACAAAGGAGCCGATCGCCAAAGGATGGGCGGACGCTTACGGTTCTAGAC 675  
 QY 655 ATCAACATCGCAAACTTGGCACTGCAAGCCGAGTTCTCGGCGGATCCACACTGCC 714  
 DB 676 ATCAACATCGCAAACTTGGCACTGCAAGCCGAGTTCTCGGCGGATCCACACTGCC 735  
 QY 715 TACTCATGCGCTACGAGCTGAGGAGTGGTCCGTAACAGCTGCGGATGGGAGCC 774  
 DB 736 TACTCATGCGCTACGAGCTGAGGAGTGGTCCGTAACAGCTGCGGATGGGAGCC 795  
 QY 775 GTCTGTGTCAGAGGCGGCAAGGCAACCAATTAGTCTCGAATTCTGCGGCGATATG 834  
 DB 796 GTCTGTGTCAGAGGCGGCAAGGCAACCAATTAGTCTCGAATTCTGCGGCGATATG 855  
 QY 835 AAACAGGTGGGTACTGATGATATAGCATGACGAGGCGGCTGTTTGAAGGCTCA 894  
 DB 856 AAACAGGTGGGTACTGATGATATAGCATGACGAGGCGGCTGTTTGAAGGCTCA 915  
 QY 895 CGACCGACACCTTACGAGCAGCAGCTTCCGCGTGACACACGCTGTTTACTGCGGTG 954  
 DB 916 CGACCGACACCTTACGAGCAGCAGCTTCCGCGTGACACACGCTGTTTACTGCGGTG 975  
 QY 955 GCGAATATGCGGCTGCGGTGCGGAAAGCTGACCTACGCGCTGACCAACGCAAGATG 1014  
 DB 976 GCGAATATGCGGCTGCGGTGCGGAAAGCTGACCTACGCGCTGACCAACGCAAGATG 1035  
 QY 1015 CCGATGTCGTCGAGCTTGGCAGCATGGTGGGCGGCGGCTGCGGCTGCAATCCGGA 1074  
 DB 1036 CCGATGTCGTCGAGCTTGGCAGCATGGTGGGCGGCGGCTGCGGCTGCAATCCGGA 1095  
 QY 1075 CTAGCCAAAGGTCTTTCGAGCAGCAAGGAGGCGTACTGTCGAAGCGGTGGCCACCGAC 1134  
 DB 1096 CTAGCCAAAGGTCTTTCGAGCAGCAAGGAGGCGTACTGTCGAAGCGGTGGCCACCGAC 1155  
 QY 1135 CTGGGGGGCGGTCACGAGCAGCAGCGGCGGCGGCTGCGGCTGACCTGCGGCTGTTAC 1194  
 DB 1156 CTGGGGGGCGGTCACGAGCAGCAGCGGCGGCGGCTGCGGCTGACCTGCGGCTGTTAC 1215  
 QY 1195 GCGGAGCAGACGTCGAGGAGTGAAGGAGGAGCGATGATGTCGCGC 1237  
 DB 1216 GCGGAGCAGACGTCGAGGAGTGAAGGAGGAGCGATGATGTCGCGC 1258  
 RESULT 3  
 V49625  
 ID V49625 standard; DNA; 1260 BP.  
 AC V49625;  
 DT 20-NOV-1998 (first entry)  
 DE Mycobacterium tuberculosis; Alanine dehydrogenase.  
 KW ss: Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.  
 OS Mycobacterium tuberculosis.  
 PN MO9836089-A2.  
 PD 20-AUG-1998.  
 PE 29-JAN-1998; E00483.  
 PR 29-JAN-1997; EP-101338.  
 PA (FLOH/) FLOHE L.  
 PI FLOHE L, Hutter B, Kolk A, Singh M;  
 WT 1; 98-457123/39.  
 PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity  
 - useful for, e.g. for diagnosis, differentiation of strains,  
 PT monitoring vaccination and identification of mycobacterial  
 PT inhibitors  
 PS Claim 13; Fig 2.3; 55pp; German.  
 CC The Mycobacterium tuberculosis alanine dehydrogenase (AladH) is in  
 CC the production of kits for diagnosing tuberculosis (TB) and other  
 CC mycobacterial infections in humans or animals. Kits are used for direct  
 CC diagnosis of TB on clinical samples (e.g. body fluids) and can  
 CC differentiate between pathogenic and non-virulent strains, e.g. for  
 CC identifying pathogens of the M. tuberculosis (M. t.) complex. The kit may  
 CC also be used to identify substances that inhibit mycobacteria, for  
 CC combatting epidemics and for vaccination follow-up. Oligonucleotides

CC derived from AladH are used similarly in diagnostic hybridisation tests,  
 CC also for culture confirmation of isolated strains and for chromosome  
 CC fingerprinting to detect/differentiate between mycobacteria, and for  
 CC L-alanine-specific biotransformation reactions. AladH is an early  
 CC antigen, present extracellularly after only a few days of growth, making  
 CC it an ideal drug target.  
 SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;  
 Query Match 98.5%; Score 1219.4; DB 1; Length 1260;  
 Best Local Similarity 99.4%; Pred. No. 1.4e-245;  
 Matches 1236; Conservative 0; Mismatches 1; Indels 6; Gaps 1;  
 QY 1 ATCTGAGATTATGAACTTCTTCATCTGAAAGCTTACATGAGAGGGGTAATC 60  
 DB 16 ATCTGAGATTATGAACTTCTTCATCTGAAAGCTTACATGAGAGGGGTAATC 75  
 QY 61 ATGCGGCTCGGTATTCGAGCAGACCAAAACAGC-----AATTCCGGTGGCCATC 114  
 DB 76 ATGCGGCTCGGTATTCGAGCAGACCAAAACAGC-----AATTCCGGTGGCCATC 135  
 QY 115 ACCCGGCGGCGTGGGAACTAACCGCTGTCGCAATGAGTCTCATCAGCAGGT 174  
 DB 136 ACCCGGCGGCGTGGGAACTAACCGCTGTCGCAATGAGTCTCATCAGCAGGT 195  
 QY 175 GCGGAGAGGCGTGGGCTATCAGCAGCAGGATTCAGAGGCGGCGGCAAGTGGTC 234  
 DB 196 GCGGAGAGGCGTGGGCTATCAGCAGCAGGATTCAGAGGCGGCGGCAAGTGGTC 255  
 QY 235 GGCACGCGCAGCAGGTGTTGGGCGGAGCTGATTTATGTCAGGTCAAGAACGATA 294  
 DB 256 GGCACGCGCAGCAGGTGTTGGGCGGAGCTGATTTATGTCAGGTCAAGAACGATA 315  
 QY 295 GCGGCGAATGAGCGCGGCGGCGGCGGCGGAGATTTGTCAGCTTCTGATTTGGCC 354  
 DB 316 GCGGCGAATGAGCGCGGCGGCGGCGGCGGAGATTTGTCAGCTTCTGATTTGGCC 375  
 QY 355 GCGTCAGTCTTTCAGCAGCAGTGGTGGTGGGAGCGGCGGCGGCGGCGGCGGAG 414  
 DB 376 GCGTCAGTCTTTCAGCAGCAGTGGTGGTGGGAGCGGCGGCGGCGGCGGCGGAG 435  
 QY 415 ACCGTCAGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 474  
 DB 436 ACCGTCAGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 495  
 QY 475 CGACTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 534  
 DB 496 CGACTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 555  
 QY 535 CTGATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 594  
 DB 556 CTGATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 615  
 QY 595 GCGGCGTACAGCAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 654  
 DB 616 GCGGCGTACAGCAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 675  
 QY 655 ATCAACATCGCAAACTTGGCACTGCAAGCCGAGTTCTCGAATTCTGCGGCGATATG 714  
 DB 676 ATCAACATCGCAAACTTGGCACTGCAAGCCGAGTTCTCGAATTCTGCGGCGATATG 735  
 QY 715 TACTCATGCGCTACGAGCTGAGGAGTGGTCCGTAACAGCTGCGGATGGGAGCC 774  
 DB 736 TACTCATGCGCTACGAGCTGAGGAGTGGTCCGTAACAGCTGCGGATGGGAGCC 795  
 QY 775 GTCTGTGTCAGAGGCGGCAAGGCAACCAATTAGTCTCGAATTCTGCGGCGATATG 834  
 DB 796 GTCTGTGTCAGAGGCGGCAAGGCAACCAATTAGTCTCGAATTCTGCGGCGATATG 855  
 QY 835 AAACAGGTGGGTACTGATGATATAGCATGACGAGGCGGCTGTTTGAAGGCTCA 894  
 DB 856 AAACAGGTGGGTACTGATGATATAGCATGACGAGGCGGCTGTTTGAAGGCTCA 915



Matches 604; Conservative 0; Mismatches 458; Indels 12; Gaps 4;

```

QY 61 ATGGCGGTGGGATTCGACCGGACCAAAACAAAGATTCGGGTGGCCATACCCCG 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 ATGAAGATGGGATTCGACCAAAACAAAGATTCGGGTGGCCATACCCCG 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 GCGGCGGTGGGAAATTAACCCGTCGTCATGAGTCTCT--CATCCAGGAGGTGCG 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GCAGGCGGTGATGAGCTGCTGCAAAAGCGGGCATGAGTGTGATGAGAGACGAAGCG 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 178 GAGAGAGGCTCGGCTATCAACGACGCGGATTCAGCGCGGACCGCGCAACTGTGCGC 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GGGGCTGJGTCGGGGGCTTCGATTCGAGATGAAAAAGCCGGCGACATGATCGTTCG 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 238 ACCGCGGACAGGCTGTGGCGCGGATTCGATTCGATTCGATTCGATTCGATTCGATTC 297
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 CGAAGCTGAGAGATGCTTGAGAGCGGAGAGATGCTTGAAAGTGAAGCGCGTGGCT 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 298 GCGGATATGCGCGCGCGGACACAGGCGAGATTCGTCAGCTTCCTTGGATTCGCGCGC 357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 CGAGATTCGCGATTCCTGCGCGCGCGGATTCCTGCGCGCGGATTCCTGCGCGG 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 358 TCACGTGCTTGCACCGCATGCGTGTGATTCGCGACCACTGCAATTCCTTACGACACC 417
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GCGGACGCGGTGCGGAAACCGGTCGTCGAGCAAAAGTGGTGGCGATCGCTTACGAGACG 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 418 GTCCACACCGCGGACCGGCACTACCCGCTGCTGCGCGCATGAGCAAGTCGCGCGTCA 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 GTGACACTGGCGAAGCGGCTGCGTCCGACCTGTCACCCGATGAGTAAAGTCGCGCGCGC 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 478 CTGCGCGCGCGGATTCGCTTACCACTGATGCAACCAACGAGGCGCGGCTGTGCTG 537
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 ATGTGCGTGAAGTTCGCGCGCGCGGCTTCTGAGAACCGCGCGGGAAGGCGATTTTG 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 538 ATGGCGGGGTGCGCGCGCTGCAACCGCGCGCGCGTGTGATGTCGCGCGCGCGCGC 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 CTGGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 598 GCGTACACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 GCGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 600
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 658 AACATGACAACTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 717
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 601 AACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 660
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 718 TCATCGCGCTAGCAGCTGAGGCGTCCGTCAAAGTCCGACCTGCTGATGCGCGCGC 777
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 TCCAACTGCTACCATATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 720
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 778 CTGGTCCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 837
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 TTGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 777
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 838 CCAGGTGCGGCTGCTGATATGATGATGATGATGATGATGATGATGATGATGATG 894
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 778 CCGGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 837
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 895 CGACGACCACTAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 954
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 838 CGACGCTAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 897
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 955 GCGAATATCCGCGCTGCGTGGCGAAGAGTGCATTCAGCGCGTGCAGCAAGCGGAGT 1014
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 898 GCGAATATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 954
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1015 CCGATATGCTGAGCTTGCAGCAATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 955 CCGATATGCTGAGCTTGCAGCAATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1014
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1075 CAGCAAGAGCTTTCGACGACGACGACGACGACGACGACGACGACGACGACGACG 1128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1015 CTGTAAAAAGGATCAACAGCTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1068
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

## RESULT 6

V5215/c  
V52155 standard; DNA: 28171 BP.  
ID V52155;  
AC V52155;  
DR 23-OCT-1998 (first entry)  
DE Streptococcus pneumoniae genome fragment SEQ ID NO:22.  
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
KW computer readable medium; vaccine; pharmaceutical composition; ds.  
OS Streptococcus pneumoniae.  
PN W09818931-A2.  
PD 07-MAY-1998.  
PF 30-OCT-1997: U19588.  
PR 31-OCT-1996: US-029960.  
PA (HUMA-) HDMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,  
PI Kunsch CA, Rosen CA;  
DR WPI: 98-272225/24.  
PT Computer-readable medium with recorded Streptococcus pneumoniae  
PT polynucleotide sequences - useful in diagnostic kits and assays, and  
PT pharmaceutical compositions and vaccines for Streptococcus  
PT pneumoniae  
PS Claim 1: Page 273-289; 1409pp; English.  
CC The present invention describes a computer readable medium which has  
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
CC on it, or a representative fragment or a sequence at least 95% identical  
CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
CC pneumoniae. The present invention also describes an isolated nucleic acid  
CC molecule encoding a homologue of any of the fragments of the S. pneumoniae  
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
CC by a process comprising: (a) screening a genomic DNA library using as a  
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
CC to 391, identifying members of the library which contain sequences  
CC that hybridize to the target sequence and isolating the nucleic acid  
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
CC from an organism, amplifying nucleic acid molecules whose nucleotide  
CC sequence is homologous to amplification primers derived from the  
CC fragment of the S. pneumoniae genome to prime the amplification and  
CC isolating the amplified sequences. The computer readable medium can be  
CC used in a computer-based system for identifying fragments of the  
CC S. pneumoniae genome of commercial importance, or expression modulating  
CC fragments of the S. pneumoniae genome. Products from the present  
CC invention can be used in diagnosis kits and assays, and pharmaceutical  
CC compositions and vaccines for S. pneumoniae.  
SQ Sequence 28171 BP: 8507 A; 5080 C; 6440 G; 8144 T;

Query Match 17.2%; Score 212.4; DB 1: Length 28171;

Best Local Similarity 53.9%; Fred. No. 5.3e-36;

Matches 533; Conservative 0; Mismatches 436; Indels 19; Gaps 4;

```

QY 61 ATGGCGGTGGGATTCGACCGGACCAAAACAAAGATTCGGGTGGCCATACCCCG 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2171 ATGTTAATCGAATTCGACCAAAACAAAGATTCGGGTGGCCATACCCCG 21652
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 GCGGCGGTGGGAAATTAACCCGTCGTCATGAGTCTCTCT--CATCCAGGAGGTGCG 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21651 CGAGGTGCTATGAGCTGATGATGATGATGATGATGATGATGATGATGATG 21592
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 181 GAGGCGTGGGCTATGACGCGCGCGGATTCAGGCGCGCGCGCGCGCGCGCGCGCGC 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21541 GAGGCGTGGGCTATGACGCGCGCGGATTCAGGCGCGCGCGCGCGCGCGCGCGCGC 21533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 241 GCGGACGAGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21531 GCTGTGTAAGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 21475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 301 GAATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21474 GAATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 21415
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Qy	361	CGTCTTGACACCGATGGTGTGTGGATTCGGGACACAGCTCAATTGCTTACGAGACCGTC	420
Db	21414	CCAAATTTAGACGATGCTATGTT-----AACGCAAAAAACAACGTAAACGTCTT	21367
Qy	421	CAGACCCCGCAGCGCGGCTACCCCTGCTTGGCCCGCATAGCGAAGTCCGCGTGACTC	480
Db	21366	CGTACCAATAGAACCAACTACCGGCTCTCGTTCGTATGATGAGTGTGACAGGTGTATG	21307
Qy	481	CGCGCCCAAGTTGGCGCTTACCACTGATGCGAACCCAAAGGGGCCGGGTGTCTATG	540
Db	21306	GCTTTTAAATTCGAGAGCTCTCTTCTTACTTAAGCAAGCTGGGGGCTGTGTGTACTT	21247
Qy	541	GGCGGGGCGCCGCGCTCGAACCAGCGAGCTGTGTATGTGGCGCGCACGCGCGC	600
Db	21246	GGTGTGTACAGAGTGTTCCTCCAAAGAAAAGTAACTATCATGTGGTGTGTCTCGTCGT	21187
Qy	601	TACAAACGACCCCGCATCTGCCAAAGCATGGCGCGCACCGCTTACGCTTCTAGACATCAC	660
Db	21186	ACACATGCTGCCCCCAATCGCCCTGTGTGTGGTCAAGTACATATTTTATGATTTATGT	21127
Qy	661	ATCGACAAATCTTGGCAACTCGACGCGAGATTCTGTGGCGCGGATCCACACTGCTACTCA	720
Db	21126	TCCAAAGGCTCTCTCAGTTCTTACAGAAAGTCTTTGGAAGTCAATTCCTTATATGCT	21067
Qy	721	TCGGCTTACGAGCTCGAGGGTGCCTCAAACTGTCGCCACTGTGTATTTGGGCGCTCTG	780
Db	21066	AATTCATTCACCAATTTAGAGCAAGTGTGAGAGATGCTGAATGTGTATTTGAGCCATTCTC	21007
Qy	781	GTGCGAGCGCACAGGACCAATTAATAGTCTCAATTCCTGTGCGCATGTGAACCA	840
Db	21006	ATCCCTGTGTGCAAAAGCACCAGGAATTGTGTACAGATGATGATGTCAAACTATGCTCA	20947
Qy	841	GGTGGCTACTGTGTGATATAGCATCTGACACAGCGCGGCGCTTTTGA--AGGCTACGA	897
Db	20946	GGCTCTGTA-TCGTTTGACGTGCTGTGTGACAAAGGTGGCGGTATGCAAAACAGCTACCGT	20888
Qy	898	CCGACACACTTCAGCACCGCAGCTTGCCTGTGCACAGAACCGCTGTTTACTGGCTGGCG	957
Db	20887	GTGACAAACGCGACGATGAACCCGCTCTATGAAAAACAGGTGTCTTCCACTATGCGTTCC	20828
Qy	958	AACATGCGCGGCTGTGTCGGAAGACGTGACCTTACGCGCTGACCAAGCGACGATGCGG	1017
Db	20827	AATATCCCTGTGTGGTGTGTGTGCACACTTCAACCATGTGGCCTTACCAAGTGTCACTTCTCT	20768
Qy	1018	TATGTCTCGAGCTTCCGACCACTGGCT	1045
Db	20767	TATATCGAAGCTTGTGCTGGCAAGGAT	20740
RESULT 7			
V74442/C			
V74442 standard; DNA; 9280 BP.			
ID	V74442:		
AC	16-MAR-1999 (first entry)		
DT	Staphylococcus aureus contig SEQ ID #131.		
BE	Computer readable medium; vaccine; S.aureus infection; immunodetection;		
KW	cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;		
KW	skin infection; surgical wound infection; scalded skin syndrome;		
KW	toxic shock syndrome; ds.		
OS	Staphylococcus aureus.		
FH	Key	Location/Qualifiers	
FT	misc_feature	841..900	
FT		/tag-a	
FT		/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc_feature	2641..2700	
FT		/tag-b	
FT		/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

[illegible]



Db 9043 AAGTAATTAGTTCATATGAGACTGTGCACTATACCAACGACCTTACCATTTGTTA 8984  
 QY 451 GCCCGATAGAGCAAGTCCCGGCTGCACTGCGCCAGGTTGACCATGTATG 510  
 Db 8983 TCACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8924  
 QY 511 CGAACCCAAAGGG 570  
 Db 8923 AACTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8864  
 QY 571 GTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 630  
 Db 8863 GTACATATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8804  
 QY 631 GCGCGACCGCTTACGCTTACGATCAACATGACAACTTGGCACTGACGCGGAG 690  
 Db 8803 GGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8744  
 QY 691 TTGTGGGGGGGATGCAACACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCT 750  
 Db 8743 TTGCTGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 8684  
 QY 751 CGTGCGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 810  
 Db 8683 CAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8624  
 QY 811 TCGAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 870  
 Db 8623 ACGAAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8564  
 QY 871 CAGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 927  
 Db 8563 CAGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 8504  
 QY 928 GTGACGACACGCTGATGATGATGATGATGATGATGATGATGATGATGATG 987  
 Db 8503 AAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8444  
 QY 988 ACCTACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1047  
 Db 8443 ACCTACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 8384  
 QY 1048 CCGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1103  
 Db 8383 ACGAAGACATGATGATGATGATGATGATGATGATGATGATGATGATG 8328  
 RESULT 8  
 T67971  
 ID T67971 standard; DNA; 1074 BP.  
 AC T67971:  
 DT 15-JUL-1997 (first entry)  
 DE H. pylori membrane protein ORF 05cp20518orf1.  
 KW Vaccine; prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;  
 KW membrane; amino acid; metabolism; ds.  
 OS Helicobacter pylori.  
 PH Key location/Qualifiers  
 FT 1..1074  
 FT /tag-  
 FT /note- "no stop codon given"  
 PN W09640893-AL.  
 PD 19-DEC-1996.  
 PF 06-JUN-1996; U09122.  
 PR 07-JUN-1995; US-487032.  
 PR 01-APR-1996; US-630405.  
 PA (ASTR) ASTRA AB.  
 PI Berghindh OT, Smith D, Mellgaard BL.  
 DR WPI; 97-052306/05.  
 PT Helicobacter pylori nucleic acid sequences and related  
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori

PT Infection, and to detect Helicobacter  
 PS Claim 27; Page 819; 148bp; English.  
 CC The present sequence encodes a Helicobacter pylori membrane  
 CC protein likely to contain four membrane spanning regions.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors.  
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the  
 CC bacterial DNA. The sequences were analysed for ORF of at least 180  
 CC nucleotides, and the predicted coding regions defined by computer  
 CC evaluation. To identify likely H. pylori antigens for vaccine  
 CC development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences  
 CC of interest, particular regions can be isolated from H. pylori by  
 CC PCR amplification for recombinant polypeptide production, e.g. in  
 CC E. coli hosts.  
 SQ Sequence 1074 BP; 302 A; 189 C; 297 G; 286 T;

Query Match 7/7%; Score 94.8; DB 1; Length 1074;

Best Local Similarity 46.1%; Pred. No. 1,2e-11; Matches 370; Conservative 0; Mismatches 417; Indels 15; Gaps 1;

QY 265 GATTATGCTCAAGTCAAGAACGATAGCGGGAATAGCGCGCTGCGACACGGG 324  
 Db 133 GATTGCTGCTCAAAATGCAAGAGCCTTAGACATGATATCCCTTCTCAAAAATA 192  
 QY 325 CAGATCTGTTACGCTTCTGCAATTTGCGCGCTGACGCTGCTGACCGATGCTTTG 384  
 Db 193 GCGACTGCTGTTATGATTTGATTTAGCTATCAAAAAGCTGTGGAATGTTAT 252  
 QY 385 GATTCCGCGACACGCTCAATGCTTACGAGACCGCTGACAGCGCGCGACCTACCC 444  
 Db 253 AATTAATAAATATCTTCTTATTTGCACTAATCAATGCGCGGCTTAAATAACCT 312  
 QY 445 CTGCTGCGCGCGATGAGGAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 496  
 Db 313 ATTTAGCGCGCTATGAGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 372  
 QY 496 -----GCTTACCACTGATGCAACCAAGGCGCGCGCTGCTGCTGCTGCTGCTG 549  
 Db 373 TTACTGCTTATGACATGATTAAGAGGCTTTATGGTAAAGGGGCTATGCTGAGGGGTTG 432  
 QY 550 CCGGCGCTGCAACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 609  
 Db 433 TCGGCTGCGCAAGAGCGCTAAATATCTCTTATTTGAGCGCGCTGCTGCTGCTGCTG 492  
 QY 610 GCGCGCATGCGCAACGCGATGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 669  
 Db 493 GCGAAGCTTAAAGCAATAGGGGCGCTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 552  
 QY 670 CTTCGCACTGAGCGCGAGTTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 729  
 Db 553 TTACAAAACACCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 612  
 QY 730 GAGCTCGAGGCTGCGCTCAAAAGTCCGACCTGCTGATTTGGGCGCTGCTGCTGCTG 789  
 Db 613 AATATCATTTCAAGCCTTAAAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 672  
 QY 790 GCGAAGGACCAAAATTTGCTGCAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 849  
 Db 673 AGCCAAACCCCTTAAAGTATTTTAAAGAGCATTTAAATATCATGACAGCAAGGGGTA 732  
 QY 850 CTGCTGATATAGCCATGACAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 909  
 Db 733 GTCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 792  
 QY 910 GACCAACCGAGGTTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 969  
 Db 793 TCTAACCCGCTGATGAGAAAGTTTGTGCTATTTAGCTGCTGCTGCTGCTGCTGCTGCT 852





FT	misc_feature	19201..19260	
FT	/tag=	k	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	/note=		
FT	misc_feature	21001..21060	
FT	/tag=	l	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	/note=		
FT	misc_feature	22801..22860	
FT	/tag=	m	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	/note=		
FT	misc_feature	24601..24660	
FT	/tag=	n	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	/note=		
FT	misc_feature	26401..26460	
FT	/tag=	o	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	/note=		
FT	misc_feature	28201..28260	
FT	/tag=	p	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	/note=		
FT	misc_feature	30001..30060	
FT	/tag=	q	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	/note=		
FT	misc_feature		
PN	Ep-786519-A2.		
PD	30-JUL-1997.	100117.	
PF	07-JAN-1997;		
PR	05-JAN-1996; US-009861.		
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Batash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,		
PI	Rosen CA;		
DR	WPI: 97-374922/35.		
PT	Polynucleotide(s) and proteins derived from Staphylococcus aureus -		
PT	stored on computer readable medium and used in the production of		
PS	anti-S.aureus vaccines		
PS	Claim 1: Page 452-469: 3271pp; English.		
CC	This sequence represents one of 5101 Staphylococcus aureus DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the S.aureus DNA sequences allows putative functions to be assigned so that protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are likely to encode antigens have been identified and these polypeptides can be used in a vaccine composition against S.aureus infection. The polypeptides can also be used in a kit for the immunodetection of S.aureus in a sample. S.aureus is implicated in numerous human diseases, including cellulitis, eyelid infections, food poisoning, osteomyelitis, skin and surgical wound infections, scaled skin syndrome, toxic shock syndrome, etc. Organisms transformed with the DNA sequences can be used for recombinant production of the polypeptides. The new DNA sequences (and their fragments) are useful as primers or probes for isolating homologues of any of the S.aureus DNA sequences contained on the computer readable medium.		
SQ	Sequence 31096 BP:	11857 A:	5243 C: 5477 G: 7488 T:

Query Match	Similarity	4.7%	Score 57.6	DB 1	Length 31096
Best Local	Similarity	44.9%	Pred. No. 0.00885		
Matches	440	Conservative	0	Mismatches	519
				Indels	21
				Gaps	5
QY	78	GACCGAGACCAAAACAAAGAAATTCGGGTGGCCATCACCCCGCGCGCGTCCGGAACT	137		
DB	10	GTCAAGAAATTTAAACAAGGTGAAGCGGTGACTTGACCTTGACACCCGAAATAATGCTAAAGTT	69		
QY	138	AACCCGTGCTGGCCATGAGGTCTCTATCCAGCAGATGCGCGGAGAGGCTCGGCTATCAC	197		
DB	70	AACGGATCAGGACACTAAAGTATTTGTTGAAAATAATGCTGGCATTTGTTCAAGATTTTC	129		
QY	198	CGACCGGATTTCAAGCGCGCGCGCCCACTGGTCGGCACCAGCGCCAGCATGATGGGC	257		
DB	130	TACAGATATGTTGAAAAA..AAGGCGCTTA---GATCGTAACCTACAGAACCAAGCATG--	185		
QY	258	CGACCGTATTTATTTGCTCAAGTCAAAAGAACCGATAGCGGCGGAATPACGCGCGCTGGC	317		
DB	185	-GAACTCATCTTTGTTATCAAAAGTAAAGAACCTCATGAAGACGAAATATCATATTTTCA	243		
QY	318	ACAGGGGAGATCTTGTCACGCTGCTTGATTTGGCCGCGCTACGTCGCTTGACCGATGC	377		
DB	244	AAGATCAATCAATTAATCTGGGCAATTTTTCATCTTACATCTTCAAAAAGAAATAGTAAAA	303		
QY	378	GTTGTTGGATTCGGCGACACACGCTCAATTTGCTTACGAGACGCTGCAGACCGCGAGCGGC	437		
DB	304	AATGCAAGAGTTGGTGTACTGACGATTAGTGTGTAACCAI--TATAAAAATGGAAA	360		
QY	438	ACTACCCCTGCTGCGCCGAGTAGCGAAGTCGCGCGTGACTGCTCGCCCGCAAGTTGGCGC	497		
DB	361	AGCAAAATTTATTTAGCGCCAAATGATGCTATAACAGATGCAACGCTCAGCAATTTATGGAGC	420		
QY	498	TTACACCTGATGGAACCCAGGGGGCGCGGTGTGCTGATGGGGGGGTGCCCGCGCT	557		
DB	421	TTACTACTCTGAAGCAACAATGGTGTCAAGGTACTTTAGTACTGTGTACATGAAAA	480		
QY	558	CGAACCGCGCCGACG-----TGCTGTGATCGCGCGCGCGCGCGCGCTACACGC	608		
DB	481	TGTGATTTACTGTTACTGATACATATGTGATTTTCGGTGGTAGTAGACAAACAATGC	540		
QY	609	AGCCGCGCATCGCCAACGGCATGGCGCGACGCTTACGGTTCTAGACATCAACATGCACA	668		
DB	541	AGCAAAATGTTGCTTGGGCACTAAATGCTAAATGATCATTTACGATTTAAAGATGACCG	600		
QY	669	ACTTGGGCAACTCGAGCGCCAGTCTGGCGCGCGGATCCACACTC---GCTACTCATCGGC	725		
DB	601	CATTAAATATCTTGAGATATGTATGCAAGAAAAGATGTCAAGTAGTCAATTAACAACCC	660		
QY	726	CTACGAGCTCGAGGGTGGCGTCAAAAGCTGCCACACTGGATGTGGGGCGCTCGTGGTGC	785		
DB	661	AGAAAATTTAGCAGAAACAATTAAGAAAGCAATGATTTATTCTTCAAAATTTTAAATTC	720		
QY	786	AGCGCCGAAGGACCCCAATTTAGTCTCGAATTCCTTGTCGGCGATATGAAACCAAGTGC	845		
DB	721	AGGTGCGAAACCGCAAAATTTGTTACTCTGATGATGGTTAAATCAATGAAAAAGTTTC	780		
QY	846	GGTACTGTGATATAGCCATGACACAGGCGCGCTGTTTGAAGGCTCAGACCGACACAC	905		
DB	781	AGTATTAATTCGATATAGCATATTGACCAAGAGTGAAGTAAACCAATTAAGCAAACTAC	840		
QY	906	CTACGACCAACCCGAGCTTCGCCGTGCAGACACGCTGTTTACTGCTGGTGGCAACATGCC	965		
DB	841	AATTTCTGATTCAGTGTATGAAGAAAGGTGTGATCTTATTTGGTGTACAAATCAAC	900		
QY	966	CGCCTCGGTGCGAAGAGCTGCACCTACGCGCTGACCAACGAGACATCCGATATGCT	1025		
DB	901	AGGACAGTCCCAACAACCTTCAACAATGTCATTAGCACAAGAAATATGATATATAT	960		
QY	1026	CGAGCTTGCCGACCATGGCT	1045		
DB	961	AGAAATTTGTGACAAAGCT	980		







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:03:42 ; Search time 341.15 Seconds

(without alignments)  
471.322 Million cell updates/sec

Title: US-09-362-485-4

Perfect score: 1237  
Sequence: 1 ATCTTCAGCATGTTATCGAC.....GGAAGCGATGATGCGCCG 1237

## Scoring table:

IDENTITY\_NOC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0  
Maximum DB seq length: 1000000Post-processing: Minimum Match 0%  
Listing first 45 summaries

## Database :

Issued\_Patents\_NA:\*  
1: /cgn2.6/ptodata/2/ina/5A.CONB.seq:\*  
2: /cgn2.6/ptodata/2/ina/5B.CONB.seq:\*  
3: /cgn2.6/ptodata/2/ina/5C.CONB.seq:\*  
4: /cgn2.6/ptodata/2/ina/5D.CONB.seq:\*  
5: /cgn2.6/ptodata/2/ina/6.CONB.seq:\*  
6: /cgn2.6/ptodata/2/ina/PTCUS.CONB.seq:\*  
7: /cgn2.6/ptodata/2/ina/Backfillseq1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47.5	3.8	1620	US-08-461-775-10	Sequence 10, Appl
2	47.6	3.8	28958	US-08-258-261B-6	Sequence 6, Appl
3	47.6	3.8	28958	US-08-456-837-6	Sequence 6, Appl
4	47.6	3.8	28958	US-08-457-342-6	Sequence 6, Appl
5	47.6	3.8	28958	US-08-457-646A-6	Sequence 6, Appl
6	47.6	3.8	28958	US-08-458-076A-6	Sequence 6, Appl
7	47.6	3.8	28958	US-08-764-233A-4	Sequence 4, Appl
8	47.6	3.8	28958	US-08-457-335A-6	Sequence 6, Appl
9	47.6	3.8	28958	US-08-729-214-6	Sequence 6, Appl
10	47.6	3.8	49377	US-08-764-233A-1	Sequence 1, Appl
11	46	3.7	2582	US-08-816-105A-2	Sequence 2, Appl
12	45.4	3.7	833	US-08-403-852D-3	Sequence 3, Appl
13	45.4	3.7	5392	US-08-403-852D-6	Sequence 1, Appl
14	42.4	3.4	1833	US-08-403-852D-6	Sequence 1, Appl
15	42.4	3.4	2185	US-08-177-508-3	Sequence 6, Appl
16	42.4	3.4	2185	US-08-265-310-3	Sequence 3, Appl
17	41.2	3.3	2414	US-08-461-775-11	Sequence 11, Appl
18	41.2	3.3	2668	US-07-642-734C-3	Sequence 3, Appl
19	41	3.3	20235	US-07-642-734C-7	Sequence 7, Appl
20	40.8	3.3	44377	US-08-804-198-1	Sequence 1, Appl
21	40.8	3.3	44377	US-08-804-198-1	Sequence 1, Appl
22	40.6	3.3	4257	US-08-690-473-1	Sequence 1, Appl
23	40.6	3.3	12001	US-08-458-568A-11	Sequence 11, Appl
24	40.2	3.2	2064	US-08-343-428-1	Sequence 1, Appl
25	40.2	3.2	11219	US-07-642-734C-1	Sequence 1, Appl
26	39.6	3.2	474	US-08-403-852D-14	Sequence 14, Appl
27	39.6	3.2	1524	US-08-461-775-10	Patent No. 5512669

28	39.6	3.2	1525	7	5229279-1	Patent No. 5229279
29	39.4	3.2	459	4	US-08-387-942C-35	Sequence 35, Appl
30	39.4	3.2	1998	4	US-08-387-942C-1	Patent No. 5212296
31	39.4	3.2	12588	4	US-08-387-942C-1	Sequence 1, Appl
32	39.2	3.2	420	1	US-08-470-179-148	Sequence 1, Appl
33	39.2	3.2	2353	6	PCT-US92-06840-1	Sequence 1, Appl
34	39	3.2	43280	3	US-08-804-227C-1	Sequence 1, Appl
35	38.8	3.1	8051	2	US-08-576-626A-2	Sequence 2, Appl
36	38.8	3.1	11219	4	US-07-642-734C-1	Sequence 1, Appl
37	38.6	3.1	1187	2	US-08-440-856A-2	Sequence 2, Appl
38	38.6	3.1	3231	2	US-08-074-121-4	Sequence 4, Appl
39	38.6	3.1	3231	6	PCT-US94-06447-4	Sequence 4, Appl
40	38.4	3.1	1215	4	US-08-947-726A-1	Sequence 1, Appl
41	38.4	3.1	2048	5	US-08-776-251-1	Sequence 1, Appl
42	38.2	3.1	2109	4	US-08-555-568B-20	Sequence 20, Appl
43	38	3.1	30001	1	US-08-125-458-1	Sequence 1, Appl
44	38	3.1	30001	3	US-08-474-933-1	Sequence 1, Appl
45	37.8	3.1	2588	3	US-08-796-414B-6	Sequence 6, Appl

## ALIGNMENTS

RESULT 1  
US-08-461-775-10  
Sequence 10, Application US/08461775  
Patent No. 5858773  
GENERAL INFORMATION:  
APPLICANT: MAZODIER, Philippe  
TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESS: Burns, Doane, Swecker & Mathis  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22315-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461.775  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/050.313  
FILING DATE: 10-MAY-1993  
APPLICATION NUMBER: FR 9011186  
FILING DATE: 10-SEP-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1620 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1620  
US-08-461-775-10

Query Match 3.8%; Score 47.6; DB 3: Length 1620;  
 Best Local Similarity 45.7%; Pred. No. 0.003;  
 Matches 321: Conservative 0; Mismatches 369; Indels 12; Gaps 4;

```

QY 114 CACCCGCGCGCGCTCGCGAAGTAAACCGTGTGGCCATGAGTCTCATCCAGGAGG 173
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 879 CACCTTCACCGCTCCACCGTATCGCGAGAGAGTGGCTCAAGCTCGACAGGCGG 938
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 174 TCGCGAGAGAGGCTCGGTATCATCCGAGGATTTCAGAGGCGGAGCGGCAACTGCT 233
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 939 TCTGAGCTGTGGGACACCGCGCGCGGTACCGTCAAGAGACGACAGCACTGCT 998
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 234 CCGGACCGCGCGCGAGGTGTGGCGGAGCTGATTATTGCTCAAGTCAAGAGAGGAT 293
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 999 GAGAGGCGGCGGCAAGCGCGAGAGTCCAGAGGCGGCTCGCCAGATCAAGAGGCGGAT 1098
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 294 AGCGGCGGATACGCGCGCGCTGCGACAGCGGCAATCTTTGACGTTCTTGCAATTGGC 353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1059 CGAGTCAAGCTCGAGTGGAGCGGAGAGTCCAGAGAGCGCGCTCGCAAGCTGGC 1118
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 354 C-GGTCAGTGTGCTGACGCGATGGGTGTTGATTCCGCGACCAAGTCAATGCTACG 412
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1119 CGCGGCGCTGTGCTGTGATCCGCGTGGCGCGCGCCAGAGTCCAGTCAAGAGAGCGCA 1178
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 413 AGACCGTCAAGACCGCGCGCGACGCACTACCCCTGCTTGCCTCGATGAGCGAA--GTGCG 470
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1179 GCACCGTGTGAGAGAGCGCATCTCCGACCGCGCGCGCTCGAGAGAGGCGGATGCTC 1238
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 471 CGGTGACATCCCGCGCGCGAGTGGCGCTTACCA---CCTGATGGGAAGCCAAAGGCGCG 527
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1239 CGGTGTGTGCTGCGCGCTGCGTCAAGGTCTTGAGCAACAACCTTGCGCGAC 1298
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 528 CGGTGTGTGATGGGCGGCGGTGCGCGCGCGTCAAGCGCGCGCTGATGATCG---585
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1299 CGCGGAGAGAGCGCGCGCTGTGCGCGTGTGCTCGCGCGCGCGCTGCGCTGCGTG 1358
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 585 ---CGCGGCGACCGCGCGCTCAAGAGCGCGCGCGTCCGCAAGCGCGCGCGCGCT 641
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1359 GATGCGCGAGAGCGCGCGCTGCGAGGCTATCATCACCAAGGCGCGGAGCTGGA 1418
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 642 TACGTTCTTACATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 701
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1419 CAAGGCGCGAGGCTTCAAGCGCGCGCGCGCGCGAGTACGCGCGCGCGCGCGCGT 1478
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 702 GATCAGCACTGCTACTATCATCGGCTTACAGCTGAGGCGTCCGCTCAAAAGTGGCGACT 761
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1479 CATGAGCGCGCTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 1538
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 762 GGTGATTGGGCGCGCTGCTGTGCCAGGCGCGCAAGCGCACCA 803
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1539 GCTCTGACGACGAGACCGCTGTGTCGAGAGCGCGCGCGA 1580
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 2  
 US-08-258-261B-6  
 ; Sequence 6, Application US/08258261B  
 ; Patent No. 5639949  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schupp, Thomas  
 ; APPLICANT: Ligon, James M.  
 ; APPLICANT: Beck, James Joseph  
 ; APPLICANT: Hill, Dwight Steven  
 ; APPLICANT: Ryals, John Andrew  
 ; APPLICANT: Gaffney, Thomas Deane  
 ; APPLICANT: Lam, Stephen Ting  
 ; APPLICANT: Hammer, Phillip E.  
 ; APPLICANT: Uknes, Scott Joseph  
 ; TITLE OF INVENTION: Genes for the syntheses of  
 ; TITLE OF INVENTION: antipathogenic substances  
 ; NUMBER OF SEQUENCES: 22  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Ciba-Geigy Corporation

```

STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,261B
FILING DATE: 08-JUN-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEtical: NO
ANTI-SENSE: NO
US-08-258-261B-6

```

Query Match 8%; Score 47.6; DB 1: Length 28958;  
 Best Local Similarity 44.1%; Pred. No. 0.0074;  
 Matches 286: Conservative 0; Mismatches 359; Indels 3; Gaps 2;

```

QY 148 GGCATAGAGTGTCTATCCAGAGAGTGGCGGAGAGGCTCGGCTATCAAGAGCGGAT 207
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25306 GCCACAGCAGCACTCAACCTCGCGGATGTGGCTATTGCTGCGCACACCGCGCGCAC 25365
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 208 TTCAGGCGGCGAGCGCGGCACTGTGCGGACCGCGAGCAGTGTGGCGCGCGTGTAT 267
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25366 TTCAGGCGGCGCGCGGCTGTGTAAGCCACACCGGAGAGCTCTCTCGGCTCGAC 25425
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 268 TTATGCTCAAGTCAAGACGAGATAGCGCGGAATACGCGCGCTGCGACAGGCGAG 327
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25426 TCGCTCGCCGAGCAAGCGCGCGCGCGAGACCGCTCTCGAGAGAGGATGGCCCTGCTG 25485
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 328 ATCTGTTCACGTTCTTGCAATTTGGCGGCTGACGTCCTTGACCGATGCTGTGGAT 387
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25486 AAGCTGTCTTCTTCTTCTGCGCAAGGCTTCAGTGGAGAGGATGGCCCTGCTGCTG 25545
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 388 TCGGCGACAGCTCAATTTGCTTACGAGACCGTCCAGACCGCGAGCGGCACTACCCCTG 447
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25546 CTGACCTCTCGCGCGCTTCTCGGCGCTCACTGGAAGCATGCGAGCGGCGCTGCTCT 25605
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 448 CTGCGCGGATGAGAGGAGTGGCGGCTGACCTGCGCGCGCGAGGTTGGCGCTTACAC -C 505
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25606 CAGCTCAGTGAAGCGCTGCGCGCTGCTGCGCGGAGAGAGGCGCGCGCTCTCGAG 25665
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 506 TGATGCAACCCAAAGGCGCGCGGCTGTGATGAGGCGGCGGCTGCGCGGCTGGAACCG 565
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25666 CGGCTCAGCTGTGACAGCGCGCGCTTTTCCGCTATGCTCTCCCTGGGCGCTCTG 25725
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 566 CCGAGG-TCTGTGATGCGCGCGCGAGCGCGCTTACAGCAGCGCGCGATGCGCAAC 624
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25726 CGCTCGCTCGCGCTGAAGCGCGCGCGCTGTGCGGCGACAGTGAAGGCGAGATCGCGCG 25785
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 625 GGCATGGGCGGACCGTTACGCTTACATCAATCAATCAATCAATCAATCAATCAATCAAT 684
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



Db 25786 GCCTTCGTCAGACGCTCTCTCCCTGAGAGAGCGGCGCCGCAATGCCGCTGCGCAGC 25845  
QY 685 GCGGAGTTTCGCGCGGCGGATCCACACTGCTACTCATCGGCGCTTACAGAGCTTCGAGGCTGCC 744  
Db 25846 AAGCGCTCACACCGCTGCGCGGCAAGGCGCATGGCGCGCTGCGAGCTCGGCGCTCC 25905  
QY 745 GTCAACGTCGCGACCTGTGATTGGGGCGCTGCTGCTGCTGCGAGGCGCC 792  
Db 25906 GACCTCAGACCTACCTGCTCCCTGGGGGAGAGGCTCTCATCGCC 25953

RESULT 3  
US-08-456-837-6  
; Sequence 6, Application US/08456837  
; Patent No. 5643774  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James M.  
; APPLICANT: Beck, James Joseph  
; APPLICANT: Hill, Dwight Steven  
; APPLICANT: Ryals, John Andrew  
; APPLICANT: Gaffney, Thomas Deane  
; APPLICANT: Lam, Stephen Ting  
; APPLICANT: Hammer, Phillip E.  
; APPLICANT: Uknes, Scott Joseph  
; TITLE OF INVENTION: Genes for the synthesis of  
; TITLE OF INVENTION: antipathogenic substances  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ciba-Geigy Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/456,837  
; FILING DATE: 01-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/457,205  
; FILING DATE: 01-JUN-1995  
; APPLICATION NUMBER: 08/258,261  
; FILING DATE: 08-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elmer, James Scott  
; REGISTRATION NUMBER: 36,129  
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8614  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ. ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28958 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOHETICAL: NO  
; ANTI-SENSE: NO  
; US-08-456-837-6

Query Match 3.8%; Score 47.6; DB 1; Length 28958;  
Best Local Similarity 44.1%; Pred. No. 0.007;  
Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

QY 148 GGCATGAGGTGCTCATCCAGGAGGCGGAGAGGGCTCGGCTTACACGAGCGGAT 207  
Db 25306 GCCCAGAGACCTTCAACCTCCGCGGATGTGGCTTATTCCTGAGCCACACCGCGCTCAC 25365  
QY 208 TTCAGGCGGCGAGCGGCACTGTGCGACCGCGGAGGAGGAGTGGGCGTACGCTGAT 267  
Db 25366 TTCAGAGACGCGCGCTCTCGTAGCCACAAACGAGAGAGAGTCTCTCGCGGCTCGAC 25425  
QY 268 TTATGCTCAAGGTCAAGACCGATAGGCGGGAATACGCGCGCTGCGACAGGCGAG 327  
Db 25426 TCCTCGCGCGGAGCAAGCGCGCGGAGACCGCTCTCTGAGAGGAGCGAAGCCAGCGC 25485  
QY 328 ATCTGTTCAGCTTCTTGATTTGGCGGCTGACGCTTGCACGAGTGGCTTGTGAT 387  
Db 25486 AAGCTGCTTCTGCTTCTTGGGCAAGCTTCGAGTGGAGAGGAGTGGCGCTTGTGCTG 25545  
QY 388 TCGGAGCAGCAGTCAATTGCTTACGAGACGCTCCAGACCGCGGCGGCGCTACCGCTG 447  
Db 25546 CTCGACTCTCGCGCGCTTTCGCGCGCTCAGCTGGAAGCATGGAGGCGGCGCTGCTCT 25605  
QY 448 CTGCGCGCGATGAGCAAGTCCGCGCTGCACTGCGCGCGGAGTTGGCGCTTACACAC--C 505  
Db 25606 CAGGTGAGTGGAGCTGCTGCTGCGCGCTGCGCGCGGAGAGGCGCGCTTCTCTGAC 25665  
QY 506 TGATGCAACCCCAAGGGGCGCGGCTGCTGATGGCGGGGTCGCGCGCTGCAACCGG 565  
Db 25666 CGGCTGAGCTCTACAGCCCGCTCTTGGCGCTATGCTCTCTGCGGCGCTTGTG 25725  
QY 566 CCGAGC-TGGTGTGATGCGCGCGCGCGCGCTTACAGCAGCAGCGCATCGCCAC 624  
Db 25726 CGCTCGCTCGGCGTAGAGCGCGCGCGCGCTGCTGCGCACAGTACAGGCGAGATCGCGCG 25785  
QY 625 GCGATGGCGCGCACCGCTTACGCTTACAGATCAACATGCAACAACTTGGCACTGAC 684  
Db 25786 GCGTTCGTCAGAGCGCTCTCTGCGAGAGCGCGCGCGCGCTTCTGCGAC 25845  
QY 685 GCGGAGTTGCGGCGCGGATTCACACTGCTCTACTATCGGCTAGAGAGTTCAGAGTGGC 744  
Db 25846 AAGCGCTTACACCGCTCCCGGCAAGGCGCATAGCGCGCATGCGGCTCGCGCGCTCC 25905  
QY 745 GTCAACGTCGCGACCTGTGATTGGGGCGCTGCTGCTGCTGCGAGGCGCC 792  
Db 25906 GACCTCAGACCTACCTGCTCCCTGGGGGAGAGGCTCTCATCGCC 25953

RESULT 4  
US-08-457-342-6  
; Sequence 6, Application US/08457342  
; Patent No. 5652898  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James M.  
; APPLICANT: Beck, James Joseph  
; APPLICANT: Hill, Dwight Steven  
; APPLICANT: Ryals, John Andrew  
; APPLICANT: Gaffney, Thomas Deane  
; APPLICANT: Lam, Stephen Ting  
; APPLICANT: Hammer, Phillip E.  
; APPLICANT: Uknes, Scott Joseph  
; TITLE OF INVENTION: Genes for the synthesis of  
; TITLE OF INVENTION: antipathogenic substances  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ciba-Geigy Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,342  
FILING DATE: 01-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/457,205  
FILING DATE: 01-JUN-1995  
APPLICATION NUMBER: 08/258,261  
FILING DATE: 08-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28958 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-457-342-6

Query Match 3.8%; Score 47.6; DB 1; Length 28958;  
Best Local Similarity 44.1%; Pred. No. 0.0074;  
Matches 286; Conservative 0; Mismatches 159; Indels 3; Gaps 2;

148 GGCATGAGTGTCTCATCCAGCAGGTGCGGAGAGGGCTCGGCTATCATCCAGCGGAT 207  
25306 GCCCAGCAGCAGCTCACCCTCGGCGATGTGGCTATTCGTGGCCACACCGCGCCAC 25365  
208 TTCAGGCGGCGAGCGCGCACTGTGTGCGACCGCGCAGCAGGTGTGGCGCAGCTGAT 267  
25366 TTCGAGCAGCGCGCGCTCTGTAGCCACACCGCGCAGAGCTCTCTCGCGCTCGAC 25425  
268 TTATGTCTCAAGTCAAGAAAGCATAGCGCGGGAATACGCGCGCTCGCACACGGGAC 327  
25426 TCCTGCGCCAGACCAAGCCCGCGAGCAGCTCTCGGAGGAGGAGGAGCCACCGGC 25485  
328 ATCTGTTCAGTCTTGTGATTTGGCGCGCTCAGCTGTGACCGATGCGTGTGAT 387  
25486 AACGTGTCTGTCTTCTTCTGTGGCAAGGCTCGCAGTGGGAAGGATGCGCTCTGCTG 25545  
388 TCCGCGCAGCAGTCAATTGCTTACAGAGCCGTCCAGACCGCGCAGCGCAGTACCCCTG 447  
25546 CTGACATCTCTGCGCGCTCTTCCGCGCTCAGCTGGAAGCATGCAAGCGCGCTGCTCT 25605  
448 CTGCGCGCGATGAGCGAAGTCCGCGGTGACCTGCGCGCGCGCGGTTGGCGTTACAC -C 505  
25606 CAGCTGAGTGTGAGCTCTGCTGCGCTCTGCGCGCGAGAGGCGCGCGCTCTCTGAC 25665  
506 TGAATGGAACCAAGGGGCGCGGTGTGCTGATGGCGGCGGTGCGCGCGTGAACCGG 565  
25666 CGGTGTGAGCTGTACAGCGCGCGCTTGTGGCGTATG TACCGTGGGCGCGCTCTG 25725  
566 CCGAGC-TGCTGTGATCGCGCGCGCGCGGCTACACCGACCGCGCGCTACCGAAC 624  
25726 CGCTGTGCGCTAGAGCGCGCGCGCGTGTGCGCGCAGTCAAGGCGGAGTGGCGCG 25785  
625 GGCATGAGCGCGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 684  
25786 GCTTCTGCGCAGCGCT 25845  
685 GCGAGTCTGTGCGCGCGAGTCAACACTGCTACTCATGCGCTTCAAGAGTGTGAGGGTGC 744  
25846 AAGCGCTACACACCGTGTGCGCGGAGAGGCGCGCATGGCGCGCGCTGCGCGCTCTC 25905

745 GTCAGAGTGTGCGCGAGTGTGATGAGGCGCGCTCTGTCAGCGCGCC 792  
25906 GACCTTCAGAGCTTACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25953

RESULT 5  
US-08-457-646A-6  
Sequence 6, Application US/08457646A  
Patent No. 5679560  
GENERAL INFORMATION:  
APPLICANT: Schupp, T.C. mas  
APPLICANT: Ligon, James M.  
APPLICANT: Beck, James Joseph  
APPLICANT: Hill, Dwight Steven  
APPLICANT: Ryals, John Andrew  
APPLICANT: Gaffney, Thomas Deane  
APPLICANT: Lam, Stephen Tiro  
APPLICANT: Hammer, Phillip E.  
APPLICANT: Umes, Scott Joseph  
TITLE OF INVENTION: Genes for the synthesis of  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,646A  
FILING DATE: 01-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/457,205  
FILING DATE: 01-JUN-1995

448 CTGCGCGCGATGAGCGAAGTCCGCGGTGACCTGCGCGCGCGGTTGGCGTTACAC -C 505  
25606 CAGCTGAGTGTGAGCTCTGCTGCGCTCTGCGCGCGAGAGGCGCGCGCTCTCTGAC 25665  
506 TGAATGGAACCAAGGGGCGCGGTGTGCTGATGGCGGCGGTGCGCGCGTGAACCGG 565  
25666 CGGTGTGAGCTGTACAGCGCGCGCTTGTGGCGTATG TACCGTGGGCGCGCTCTG 25725  
566 CCGAGC-TGCTGTGATCGCGCGCGCGCGGCTACACCGACCGCGCGCTACCGAAC 624  
25726 CGCTGTGCGCTAGAGCGCGCGCGCGTGTGCGCGCAGTCAAGGCGGAGTGGCGCG 25785  
625 GGCATGAGCGCGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 684  
25786 GCTTCTGCGCAGCGCT 25845  
685 GCGAGTCTGTGCGCGCGAGTCAACACTGCTACTCATGCGCTTCAAGAGTGTGAGGGTGC 744  
25846 AAGCGCTACACACCGTGTGCGCGGAGAGGCGCGCATGGCGCGCGCTGCGCGCTCTC 25905

Query Match 3.8%; Score 47.6; DB 1; Length 28958;  
Best Local Similarity 44.1%; Pred. No. 0.0074;  
Matches 286; Conservative 0; Mismatches 159; Indels 3; Gaps 2;

148 GGCATGAGTGTCTCATCCAGCAGGTGCGGAGAGGGCTCGGCTATCATCCAGCGGAT 207  
25306 GCCCAGCAGCAGCTCACCCTCGGCGATGTGGCTATTCGTGGCCACACCGCGCCAC 25365  
208 TTCAGGCGGCGAGCGCGCACTGTGTGCGACCGCGCAGCAGGTGTGGCGCAGCTGAT 267  
25366 TTCGAGCAGCGCGCGCTCTGTAGCCACACCGCGCAGAGCTCTCTCGCGCTCGAC 25425

QY 268 TTATTGCTCAGAGTCAAGAACCCGATAGCGGGGAAATACGGCCCTCGGACACGGGCGAG 327  
 Db 25426 TCGCTCGCCGACGACGACCCGCGGAGACCGCTCCCTCGGACGAGGAGCCACGGC 25485  
 QY 328 ATCTTGTCAGCTTTGCTCATTTGGCCGGGACGCTGCTGGACGAGTGGCTGTGGAT 387  
 Db 25486 AAGCTGCTTGTGCTTTTCTTCCGAGGAGAGCTGCAAGTGGAGAAAGGATGCCCTCTGG 25445  
 QY 388 TCCGGGACACGTCATTTGCTTACGAGACCGCTCCAGACCGCGGACGAGGACATCCCTG 447  
 Db 25446 CTCGACTCTCGCCGCTTTCGCGCTCAGTCAAGCATGGAGCGGGCGCTGCTCCT 25605  
 QY 448 CTGGCCCGGAGAGGAGAGTGGCGGCTGAGCTCGCGCCGCGGAGTGGCGCTTACAC--C 505  
 Db 25606 CAGCTGAGAGTGGAGCTCTCCCTCTCTGCGCCGCGGAGAGGGGCGCCCTCTGAC 25665  
 QY 506 TGATCGAACCCAGAGGGGCGCGGCTGTGCTGATGGGGGGTGGCGGCGCTGCAACCGG 565  
 Db 25666 CGCGTCGACGCTGATACAGCCGCGCTTTTGGCGCTCATGCTTCCCTGGCGGCTCTGG 25725  
 QY 566 CCGAGC-TGCTGTGATGCGGCGCGGACCGCGGCTACACGACCGCGATCGCAA 624  
 Db 25726 CGCTGCTCGGCGTAGAGCCGCGCGCTGCTGCGGACAGTCAAGGAGATCGCGCGC 25785  
 QY 625 GGCATGGGCGGACCGCTTACGCTTACATCAATCAATCGAACAATTGCGCACTCGAC 684  
 Db 25786 GCTTGTGCGAGGCGGCTCTCTCTCTGAGAGAGGCGCGCGGCGCGCTGCGGCGAGC 25845  
 QY 685 GCCGAGTCTGCGGCGGATCCACATCGCTACTCATCGGCTTACGAGTGGGATGCGC 744  
 Db 25846 AAGGCGCTACACCGTGGCGGCAAGGGGCGCATAGCGCGCGCTGAGCTGCGCGCTTC 25905  
 QY 745 GTCAACGTCGCCGACCTGTGATTGGGCGCTCTGTCGACAGGCGC 792  
 Db 25906 GACCTCGACCTACTCTGCTCTGCGGCGACAGGCTCTCATGCGC 25953

RESULT 6  
 US-08-458-076A-6  
 Sequence 6, Application US/08458076A  
 Patent No. 5698425  
 GENERAL INFORMATION:  
 APPLICANT: Schnupp, Thomas  
 APPLICANT: Ligon, James M.  
 APPLICANT: Beck, James Joseph  
 APPLICANT: Hill, Dwight Steven  
 APPLICANT: Ryals, John Andrew  
 APPLICANT: Gaffney, Thomas Deane  
 APPLICANT: Lam, Stephen Ting  
 APPLICANT: Hammer, Phillip E.  
 APPLICANT: Uknes, Scott Joseph  
 TITLE OF INVENTION: Genes for the synthesis of  
 antipathogenic substances  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ciba-Geigy Corporation  
 STREET: 7 Skyline Drive  
 CITY: Haverl Jorne  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10532  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/458, 076A  
 FILING DATE: 01-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/457, 205

FILING DATE: 01-JUN-1995  
 APPLICATION NUMBER: 08/258,261  
 FILING DATE: 08-Jun-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Elmer, James Scott  
 REGISTRATION NUMBER: 36,129  
 REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-541-8614  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SRO ID NO: 6  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 28958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-458-076A-6

Query Match 3.8%; Score 47.6; DB 2: Length 28958;  
 Best Local Similarity 44.18; Pred. No. 0.0074;  
 Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

QY 148 GGCATGAGAGTGCATTCACAGCAGTGGCGGAGAGGGGCTGCTATACCGAGCGGAT 207  
 Db 25306 GCCACAGCAGACCTCACCCTCGCGGATGTGGCTATTGCTGCGCACACCGCGCCAC 25365  
 QY 208 TTCACGCGGCGGCGCACTGTGCGGACCGCGGACGAGTGTGGCGGAGTGTAT 267  
 Db 25366 TTCGACGCGCGCTCTCTGATGACCCAGCAGGAGCTCTCTCGGGCTGGAC 25425  
 QY 268 TTATTGCTCAGAGTCAAGAACCCGATAGCGGGGAAATACGGCCCTCGGACACGGGCGAG 327  
 Db 25426 TCGCTCGCCGACGACGACCCGCGGAGACCGCTCCCTCGGACGAGGAGCCACGGC 25485  
 QY 328 ATCTTGTCAGCTTTGCTCATTTGGCCGGGACGCTGCTGGACGAGTGGCTGTGGAT 387  
 Db 25486 AAGCTGCTTGTGCTTTTCTTCCGAGGAGAGCTGCAAGTGGAGAAAGGATGCCCTCTGG 25445  
 QY 388 TCCGGGACACGTCATTTGCTTACGAGACCGCTCCAGACCGCGGACGAGGACATCCCTG 447  
 Db 25446 CTCGACTCTCGCCGCTTTCGCGCTCAGTCAAGCATGGAGCGGGCGCTGCTCCT 25605  
 QY 448 CTGGCCCGGAGAGGAGTGGCGGCTGAGCTCGCGCCGCGGAGTGGCGCTTACAC--C 505  
 Db 25606 CAGCTGAGAGTGGAGCTCTCCCTCTCTGCGCCGCGGAGAGGGGCGCCCTCTGAC 25665  
 QY 506 TGATCGAACCCAGAGGGGCGCGGCTGTGCTGATGGGGGGTGGCGGCGCTGCAACCGG 565  
 Db 25666 CGCGTCGACGCTGATACAGCCGCGCTTTTGGCGCTCATGCTTCCCTGGCGGCTCTGG 25725  
 QY 566 CCGAGC-TGCTGTGATGCGGCGCGGACCGCGGCTACACGAGCGCCATCGCAA 624  
 Db 25726 CGCTGCTCGGCGTAGAGCCGCGCGCTGCTGCGGACAGTCAAGGAGATCGCGCGC 25785  
 QY 625 GGCATGGGCGGACCGTTCAGGCTTACGATCAATCAATCGAACAATTGCGCAACTGAC 684  
 Db 25786 GCTTGTGCGAGGCGGCTCTCTCTGAGAGAGCGCGCGGCGGCTCTGCGGAGC 25845  
 QY 685 GCCGAGTCTGCGGCGGATCCACATCGCTACTCATCGGCTTACGAGTGGGATGCGC 744  
 Db 25846 AAGGCGCTACACCGTGGCGGCAAGGGGCGCATAGCGCGCGCTGAGCTGCGCGCTTC 25905  
 QY 745 GTCAACGTCGCCGACCTGTGATTGGGCGCTCTGTCGACAGGCGC 792  
 Db 25906 GACCTCGACCTACTCTGCTCTGCGGCGACAGGCTCTCATGCGC 25953

RESULT 7  
 US-08-764-233A-4

```

: Sequence 4 Application US/08764233A
: Patent No. 5716849
:
: GENERAL INFORMATION:
: APPLICANT: Ligon, James M.
: APPLICANT: Schupp, Thomas
: APPLICANT: Beck, James J.
: APPLICANT: Hill, Dwight S.
: APPLICANT: Neff, Snezana
: APPLICANT: Ryals, John A.
: TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 520 White Plains Road, P.O. Box 2005
: CITY: Tarrytown
: STATE: NY
: COUNTRY: USA
: ZIP: 10591
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/764,233A
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/729,214
: FILING DATE: 09-OCT-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/258,261
: FILING DATE: 08-JUN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Melgs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: 1506/CIP6
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (919) 541-8587
: TELEFAX: (919) 541-8689
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Sorangium cellulosum
: IMMEDIATE SOURCE:
: CLONE: p98/1
:
: US-08-764-233A-4

```

	Query Match	Best Local Similarity	Score 3.8%	DB 47.6	Length 28958
	Matches 286	Conservative	0	Mismatches 359	Indels 3
				Gaps 2	
OY	148	GCCCATGAGTGTCTATCCAGCAGCGTGC	CGGAGAGGGCTGGCTATACCGACCGCGAT	207	
Db	25306	GCCCATGAGTGTCTATCCAGCAGCGTGC	CGGAGAGGGCTGGCTATACCGACCGCGAT	25365	
OY	208	TTCAAGCGCGCAGCGACGCGACGTCGTCG	CGGAGAGGGCTGGCTATACCGACCGCGAT	267	
Db	25366	TTCAAGCGCGCAGCGACGCGACGTCGTCG	CGGAGAGGGCTGGCTATACCGACCGCGAT	25425	
OY	268	TTATTCGCTCAAGTCAAGAACCGATAGCG	CGGAGAGGGCTGGCTATACCGACCGCGAT	327	
Db	25426	TTATTCGCTCAAGTCAAGAACCGATAGCG	CGGAGAGGGCTGGCTATACCGACCGCGAT	25485	
OY	328	ATCTGTTGTCACGTTTTCGATTGGCGCGGT	AGAGTGTTCACCGACCGCGAT	387	
Db	25486	ATCTGTTGTCACGTTTTCGATTGGCGCGGT	AGAGTGTTCACCGACCGCGAT	25545	

QY	388	TCGCGGCAACAGTCATTTGCTTACAGAGACCGTCCAGACCGGCGAAGGCGGACATACCCG	447
Db	25546	CTGAGCTCTCTGGCCGCTCTTCGGCCCTAGCTCGAAGATATCGAGCGCGGCTTGCTCCT	25605
QY	448	CTTGCCCGATGAGGAGAGTCGCGCGGCTGAGTCCGCGCGCCAGTTGGCTGTTACCAAC--C	505
Db	25606	CACGTGAGATGAGAGCTGCTGCTCCGCTCTGCGCGCGGACGAGGGCGCCCTCCTCCGTGAC	25665
QY	506	TGATGCCAACCCAGAGGGGCGCCGGTGTGCTGATGAGGGGGGATGCCGCGTTCGAACCG	565
Db	25666	CGCGTCACAGCTGTACAGGCCGCCCTCTTCCGTCATGTTGTTCTCCCTGGCGGCTCTG	25725
QY	566	CCGAGC--TCGTGGTGTATCGGCGCCGCGACCGCGGCTTAAACGAGCGCCGATCGGCACAC	624
Db	25726	CGCTCGCTGGGCTAAAGCCCCCGCCCTCTGTCGGCCACAGTCAGGGCGAGATCGCCGCG	25785
QY	625	GGCATGGGCGGACCGTTACGGTTCTAGACATCAACATCGACAACTTTCGGCAACTCGAC	684
Db	25786	GCTTCGTGTCGACAGGCGCTCTCTCCCTCGAGAGCGGGCGCGATCGCGCGCCCTGGCCAC	25845
QY	685	GCGGAGTTCTGGGGCGGATCCACACTGCTACTCATGGCTTACGAGCTCGAGGGTGCC	744
Db	25846	AAACCGGTCAACCGCTCCGCCGCGAACGGGGCCATATGGCCCGCTCGAGCTCGGCGGCTCC	25905
QY	745	GTCAAACGTGCGAGCTGGTGATTTGGGGCGCTCTCGTGGCAGGCGGC	792
Db	25906	GACCTCCAGACTACTCTGCTCTCCCTGGGGGCGACAGGCTTTCATTCGC	25953

RESULT 8  
 US-08-457-335A-6  
 Sequence 6, Application US/08457335A  
 Patent No. 5723759  
 GENERAL INFORMATION:  
 APPLICANT: Schnupp, Thomas  
 APPLICANT: Ligon, James M.  
 APPLICANT: Beck, James Joseph  
 APPLICANT: Hill, Dwight Steven  
 APPLICANT: Ryals, John Andrew  
 APPLICANT: Gaffney, Thomas Deane  
 APPLICANT: Iam, Stephen Ting  
 APPLICANT: Hammer, Phillip E.  
 APPLICANT: Uknes, Scott Joseph  
 TITLE OF INVENTION: Genes for the synthesis of  
 NUMBER OF INVENTION: antipathogenic substances  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ciba-Geigy Corporation  
 STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10532  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/457,335A  
 FILING DATE: 01-JUN-1995  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/457,205  
 FILING DATE: 01-JUN-1995  
 APPLICATION NUMBER: 08/258,261  
 FILING DATE: 08-Jun-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Elmer James Scott  
 REGISTRATION NUMBER: 36,129  
 REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
 TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8614  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 28958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 US-08-457-335A-6

Query Match 3.8%; Score 47.6; DB 2; Length 28958;  
 Best Local Similarity 44.1%; Pred. No. 0.0074;  
 Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

QY 148 GGCAGTATGAGTCTATCATCCAGCAGGTGCGGAGAGGGCTATCATCCAGCGGGAT 207  
 DB 25306 GCCCAGCAGCAGCTCAGCCCTCGGGATGTGCTTATCGCTGACACACCCCGCCAC 25365  
 QY 208 TTCAAGCGCGCAGGCGGCACTGATGCGACCGCGGAGGAGGAGGAGGAGGAGGAG 267  
 DB 25366 TTGAGAGCAGCGCGCGCTCTCGTAGCCACACCGGAGAGCTCTCTCGCGCTCGAC 25425  
 QY 268 TTATTCCTCAGGTCAAGAAAGCAGATAGCGGCGAAATACGCGCGCTCGACACAGGGCAG 327  
 DB 25426 TCGCTCGCCAGGAGACCGCGCGCGAGCAGCGTCTCTCGAGCGAGGAGGAGCAGCGG 25485  
 QY 328 ATCTTGTACGTTCTTGTGATTTGGCGCGCTCAGAGCTTGCACCGATGCTTTGGAT 387  
 DB 25486 AAGCTGCTCTTCTCTCTCTCTGAGGAGGCTGCAAGGAGGAGGAGGAGGAGGAGG 25545  
 QY 388 TCCGAGCAGCAGTCAATTCGCTACAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 447  
 DB 25546 CTGAGCT 25605  
 QY 448 CTTCGCCGATGAGCAGAGTCCGCGTGCAGTCCGCGCGCGCGCGCGCGCGCGCGCG 505  
 DB 25606 CAGCTGAGAGTGGAGCTGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25665  
 QY 506 TGAATGAGACCCAAAGGGGGGGGGGGGGTGTGTGATGGGGGGGGGGGGGGGGGGGG 565  
 DB 25666 CGGCTGAGCTCTTACAGCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25725  
 QY 566 CCGAGC-TGCTGTGAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 624  
 DB 25726 CGCTCGCTGGCGCTAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25785  
 QY 625 GCGATGGCGCGCGCGCTTACAGCTTACAGATACATGACAACTTGGGCAACTCGAC 684  
 DB 25786 GCGTGTGAGCAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25845  
 QY 685 GCCGATGCTGCG 744  
 DB 25846 AAGGCGCTCAGCAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25905  
 QY 745 GTCAACAGTCCGAGCTGTGATTTGGGGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 792  
 DB 25906 GACCTCAAGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25953

## RESULT 9

US-08-729-214-6

Sequence 6, Application US/08729214

Patent No. 5817502

## GENERAL INFORMATION:

APPLICANT: Ligon, James M.  
 APPLICANT: Hill, Dwight Steven  
 APPLICANT: Ryals, John Andrew  
 APPLICANT: Hammer, Phillip E.  
 APPLICANT: van Pee, Karl-Heinz

APPLICANT: Kliner, Sabine  
 TITLE OF INVENTION: Genes for the synthesis of  
 TITLE OF INVENTION: antipathogenic substances  
 NUMBER OF SEQUENCES: 27  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: Ciba-Geigy Corporation  
 STREET: 520 White Plains Road  
 CITY: Tarrytown  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10591  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/729,214  
 FILING DATE: TBA  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meigs, J. Timothy  
 REGISTRATION NUMBER: 38,241  
 REFERENCE/DOCKET NUMBER: CGC 1506/CIP5  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-541-8587  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 28958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 US-08-729-214-6

Query Match 3.8%; Score 47.6; DB 2; Length 28958;  
 Best Local Similarity 44.1%; Pred. No. 0.0074;  
 Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

QY 148 GGCAGTATGAGTCTATCATCCAGCAGGTGCGGAGAGGGCTATCATCCAGCGGGAT 207  
 DB 25306 GCCCAGCAGCAGCTCAGCCCTCGGGATGTGCTTATCGCTGACACACCCCGCCAC 25365  
 QY 208 TTCAAGCGCGCAGGCGGCACTGATGCGACCGCGCGCGCGCGCGCGCGCGCGCG 267  
 DB 25366 TTGAGAGCAGCGCGCGCTCTCGTAGCCACACCGGAGAGGAGGAGGAGGAGGAG 25425  
 QY 268 TTATTCCTCAGGTCAAGAAAGCAGATAGCGGCGAAATACGCGCGCTCGACACAGGGCAG 327  
 DB 25426 TCGCTCGCCAGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 387  
 QY 328 ATCTTGTACGTTCTTGTGATTTGGCGCGCTCAGAGCTTGCACCGATGCTTTGGAT 387  
 DB 25486 AAGCTGCTCTTCT 25545  
 QY 388 TCCGAGCAGCAGTCAATTCGCTACAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 447  
 DB 25546 CTGAGCT 25605  
 QY 448 CTTCGCCGATGAGCAGAGTCCGCGTGCAGTCCGCGCGCGCGCGCGCGCGCGCGCG 505  
 DB 25606 CAGCTGAGAGTGGAGCTGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25665  
 QY 506 TGAATGAGACCCAAAGGGGGGGGGGGTGTGTGATGGGGGGGGGGGGGGGGGGGG 565  
 DB 25666 CGGCTGAGCTCTTACAGCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25725  
 QY 566 CCGAGC-TGCTGTGAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 624

DB 25726 CGCTGCTGCGGGGTAGAGCCGCGCGCTGCTGCGGCACAGTCAGGCGGAGATCGCCGCC 25785  
QY 625 GGCATGGGCGCGACGCTTACGTTAGACATCAACATGACAAACTTGGGACATCGAC 684  
DB 25786 GCGTTGCTGCGAGCGCTCTCTCCCTCGAGAGCGGCGCCGATCGCCCGCTCGCCAGC 25845  
QY 685 GCGGATCTGCGGCGCGGATTCACACTGCTACTCATCGGCTTACAGAGCTGAGGCTGCC 744  
DB 25846 AAGGCGCTACACGCTGCGCGGACAGGCGGCGGCGGCGGCGGCTGAGCTCGGCGCTCC 25905  
QY 745 GTCMAACGTGCGGACCTGATGAGTGGGCGGCTGCTGCTGCTGCGGCGGCC 792  
DB 25906 GACCTTCAGACCTACTCTGCTGCGGCGGACAGGCTCTCATCGCC 25953

RESULT 10  
US-08-764-233A-1  
Sequence 1, Application US/08764233A  
Patent No. 5716849  
GENERAL INFORMATION:  
APPLICANT: Ligou, James M.  
APPLICANT: Schupp, Thomas  
APPLICANT: Beck, James J.  
APPLICANT: Hill, Dwight S.  
APPLICANT: Neff, Suzanna  
APPLICANT: Ryals, John A.  
TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 520 White Plains Road, P.O. Box 2005  
CITY: Tarrytown  
STATE: NY  
COUNTRY: USA  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764,233A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/729,214  
FILING DATE: 09-06-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/258,261  
FILING DATE: 08-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: 1506/CIP6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 541-8587  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 49377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Sorangium cellulosum  
IMMEDIATE SOURCE:  
CLONE: p98/1, pUL3, and pVKM15  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 383..760  
OTHER INFORMATION: /product= "SorB"  
OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous

OTHER INFORMATION: the reductase domains of type I PKs such as eryA from  
OTHER INFORMATION: Saccharopolyspora erythraea."  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 927..19874  
OTHER INFORMATION: /product= "SorA"  
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs t  
OTHER INFORMATION: are known to be involved in the synthesis of polyketide  
OTHER INFORMATION: compounds."  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 942..7115  
OTHER INFORMATION: /product= "Module 1 of SorA"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 7203..12884  
OTHER INFORMATION: /product= "Module 2 of SorA"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 13455..19616  
OTHER INFORMATION: /product= "Module 3 of SorA"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 19871..46318  
OTHER INFORMATION: /product= "SorB"  
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs ge  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 19870..24556  
OTHER INFORMATION: /product= "Module 1 of SorB"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 24638..30820  
OTHER INFORMATION: /product= "Module 2 of SorB"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 30881..35446  
OTHER INFORMATION: /product= "Module 3 of SorB"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 35528..40114  
OTHER INFORMATION: /product= "Module 4 of SorB"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 40190..46318  
OTHER INFORMATION: /product= "Module 5 of SorB"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 46851..47891  
OTHER INFORMATION: /product= "SorM"  
OTHER INFORMATION: /note= "The protein encoded by the sorM gene is highly  
OTHER INFORMATION: homologous to the methyltransferase from Streptomyces  
OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the  
OTHER INFORMATION: polyketide rapamycin."  
US-08-764-233A-1

Query Match 3 84; Score 47.6; DB: 2; Length 49377;  
Best Local Similarity 44.18; Pred. No. 0.0086;  
Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

QY 148 GGCATGAGTGTCTATTCAGGCGAGTGGCGGAGAGGCTGCTATACCGAGCGGAT 207  
DB 41609 GCCACGACGACCTACCTCGCGGATGTGGCTATTCGCTGCGACACCGCGCCAC 41668  
QY 208 TTCAAGGCGGAGCGGCGCACTGTGCGACCGCGGACGAGTGTGGCGGAGCTGAT 267  
DB 41669 TTCGAGCAGCGCGCGCTCTGTAGCCACACGCGGACGAGCTCTGCGGCTGAC 41728  
QY 268 TTATGCTCAAGTCAAGAACGATAGCGGCGGAATACGCGCGCTTGCACACGGGCGAG 327  
DB 41729 TCGCTGCGCGGACGAGCGCGCGCGCGGAGCAGCGCTCTGAGAGGAGGAGAACCGAGCGC 41788

QY 328 ATCTTGTTCACGTTCTTCTGCAATTTGGCCGCTCAGTGTCTGACGAGATGCTTGTGTGAT 387  
 Db 41789 AAGCTGTCTTGTCTTCTTCTTCTGAGAGGCTGCAAGTGGGATGGCCCTCTGCTG 41848  
 QY 388 TCCGACACCACTCAATTTGCTTCTGAGAGCCGCTCCAGACCCGCGACGCGCTACCCCTG 447  
 Db 41849 CTCGACTCTCTCCGCTCTTCCGCGCTCAGCTCGAAGCTGCGAGCGCCGCTGCTCT 41908  
 QY 448 CTTGGCCCGATGAGGAGTGTCCGCGTGCAGTGGCGCCGCTGAGGCTTACACAC--C 505  
 Db 41909 CACGTGTGAGTGAAGCTGTCTGCGCTGCTGCGCCGCTGAGAGAGGCGCCCTCTCTGAC 41968  
 QY 506 TGATGCGAAGCCAGAGGCGCGCGGTGTGCTGATGAGGCGGCTGCGCGCTGAGAACCG 565  
 Db 41969 CGCGTGTGAGTGTGATGAGCGCCCTCTTGTGCGGTGATGTGTCTCTGCGCGCGCTG 42028  
 QY 566 CCGAGC-TGTGTGTGCGCGCGCGCGCGCGCGCTTACAGAGCAGCGCGCTGCGCAAC 624  
 Db 42028 CCGTGTGCGCGGTGAGAGCGCGCGCGCGCGCTGCGCGCAAGTGAAGCGAGTCCGCG 42088  
 QY 625 GGCATGCGCGCGCGCGCTTACGCTTCTGATGAGTCAATCAATGAGTGTGCGCAACTGAC 684  
 Db 42089 GCTTGTGTGCGCGCGCTTCTCTCTGAGAGCGCGCGCGCTGCGCGCTGCGCGAGC 42148  
 QY 685 GCGGAGTGTGCGCGCGCGCTTCTCTCTGAGAGCGCGCGCGCTGCGCGCTGCGCGAGC 744  
 Db 42149 AAGGCGCTGAGAGCGCGCGCGCGCGCGCGCGCTGCGCGCTGCGCGAGTGTGCGCG 42208  
 QY 745 GTCAAGCGTGTGCGCGCTGCTGATGTGCGCGCGCTGCTGCGCGCGCGCGCGCGCG 792  
 Db 42209 GACTGTGAGAGCTGCTGCTGCTGCGCGCGCGCGCGCGCGCTGCTGCGCGCGCG 42256

## RESULT 11

US-08-816-105A-2  
 ; Sequence 2, Application US/08816105A

; Patent No. 589882

; GENERAL INFORMATION:

; APPLICANT: Crueger, Anneliese; Dellweg, Hans-Georg; Lenz,

; APPLICANT: Jurgen; Schroder, Werner; Pape, Hermann;

; APPLICANT: Gecke, Klaus; Schaper, Beate; Hemker, Michael;

; APPLICANT: Plepersberg, Wolfgang; Distler, Jurgen;

; APPLICANT: Stralman, Ansgar

; TITLE OF INVENTION: PROCESSES FOR PREPARING ACARVIOSTYL

; TITLE OF INVENTION: TRANSFERASE AND FOR USING IT IN THE

; TITLE OF INVENTION: CONVERSION OF ACARBOSE HOMOLOGUES

; TITLE OF INVENTION: INTO ACARBOSE, FOR THE PREPARATION

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sprung Kramer Schaefer & Briscoe

; STREET: 660 White Plains Road

; CITY: Tarrytown

; STATE: New York

; COUNTRY: USA

; ZIP: 10591-5144

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage

; OPERATING SYSTEM: System 7.5

; SOFTWARE: Wordperfect 3.5

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/816,105A

; FILING DATE: 14-MAR-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: DE 19625269.5

; FILING DATE: 25-JUN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: DE 19611252.4

; FILING DATE: 22-MAR-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Kuitl G. Briscoe

; REGISTRATION NUMBER: 33,141  
 ; REFERENCE/DOCKET NUMBER: Bayer 9814-KGB  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (914) 332-1700  
 ; TELEFAX: (914) 332-1844  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2582 nucleotides  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-816-105A-2

Query Match 3.7%; Score 46; DB 4; Length 2582;  
 Best Local Similarity 47.3%; Pred. No. 0.0088;  
 Matches 139; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 353 CCGCGTACGCTCTGACAGTACGCTTGTGATTCGCGACACGATTCATTCCTACG 412  
 Db 1576 CCGCGTACG 1635  
 QY 413 AACCGTCCAGACCG 472  
 Db 1636 TGACCGTCCAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1695  
 QY 473 GTCAAGTGTGCG 532  
 Db 1696 GCGAGTGTGCG 1755  
 QY 533 TCTGTATGCG 592  
 Db 1756 TCGCGCGCGTGTGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1815  
 QY 593 CCGCGCGTCAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 646  
 Db 1816 CCGCGCGAGACCGTGTGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1869

## RESULT 12

US-08-403-852D-3  
 ; Sequence 3, Application US/08403852D

; Patent No. 5891695

; GENERAL INFORMATION:

; APPLICANT: Blanc, Veronique

; APPLICANT: Bianchi, Francis

; APPLICANT: Crozet, Joel

; APPLICANT: Jacques, Nathalie

; APPLICANT: Lacroix, Patricia

; APPLICANT: Thibaut, Denis

; APPLICANT: Zagorec, Monique

; APPLICANT: Debussche, Laurent

; APPLICANT: De Crecy-Lagard, Valerie

; TITLE OF INVENTION: Polypeptides involved in the

; TITLE OF INVENTION: Biosynthesis of Streptogramins, Nucleotide Sequences

; NUMBER OF SEQUENCES: 43

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Flanagan, Henderson, Farabow, Garrett & Dunner

; STREET: 1300 I Street, N.W., Suite 700

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/403,852D

; FILING DATE: 10-MAY-1995

; PRIOR APPLICATION DATA:



Page 10

RESULT 14  
US-08-403-852D-6  
; Sequence 6, Application US/08403852D  
; Patent No. 581655  
; GENERAL INFORMATION:  
; APPLICANT: Blanc, Veronique  
; APPLICANT: Blanche, Francis



Db 1310 CCCCAGCGGCTTCTTCTACACCGAGGCCGTGCTCGGCTTCATGTGGGGCCGAGCA 1369  
QY 645 GGTTCATAGCATCAACATCGACAACCTTCGCAACTCGACCGAGTTCTGCGCCGAT 704  
Db 1370 CGGGGTGAGCTGACCAACAACGCTATTACCCGACCCGTGGTACTTCAACTGCAAGGA 1429  
QY 705 CCACACTCGCTACTCATCGGCTACGAGCTCGAGGGTCCCTCAACGCTGCCGACTGGT 764  
Db 1430 CGACCCCGACAGAAAGCGCTCGTCGAGGCCGTCTCGGGGCTCCCGGTACGCGGAGAA 1489  
QY 765 GATTGGGGCCGCTCCTGCTGCGCAGGCGCC 792  
Db 1490 GAAGGGCGCGGTCAACGCTCGCCGGGCGCC 1517

Search completed: June 22, 2000, 15:06:11  
Job time: 17849 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:07:22 ; Search time 5541.94 Seconds  
(Without alignments)  
904.708 Million cell updates/sec

Title: US-09-362-485-4  
Perfect score: 1237  
Sequence: 1 ATCTGCGATTTATCGAAC.....CGAGCGATGATCGGCCG 1237

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 segs, 2026611650 residues  
Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_est48:\*  
83: gb\_est49:\*  
84: gb\_est50:\*  
85: gb\_est51:\*  
86: em\_est52:\*  
87: em\_est53:\*  
88: em\_est54:\*  
89: em\_est55:\*  
90: gb\_est56:\*  
91: gb\_est57:\*  
92: gb\_est58:\*  
93: gb\_est59:\*  
94: gb\_est60:\*  
95: em\_est61:\*  
96: em\_est62:\*  
97: em\_est63:\*  
98: em\_est64:\*  
99: em\_est65:\*  
100: em\_est66:\*  
101: em\_est67:\*  
102: gb\_est68:\*  
103: gb\_est69:\*  
104: em\_est70:\*  
105: gb\_est71:\*  
106: gb\_est72:\*  
107: gb\_est73:\*  
108: gb\_est74:\*  
109: gb\_est75:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES

No.	Score	Match	Length	DB	ID	Description
1	57.4	4.6	925	82	CNS0091P	AL053013 Drosophila
2	55.2	4.5	925	82	CNS0091P	AL053013 Drosophila
3	51.8	4.2	535	74	AV390505	AL053013 Drosophila
4	50.2	4.1	844	82	CNS0052P	AL056652 Drosophila
5	49.0	4.0	910	82	CNS0060N	AL056652 Drosophila
6	46.6	3.8	846	82	CNS010R3	AL059337 Drosophila
7	46.6	3.8	1101	82	CNS0181E	AL108764 Drosophila
8	46.4	3.8	935	82	CNS006KX	AL068051 Drosophila
9	46.2	3.7	645	82	CNS01213	AL101589 Drosophila
10	45.8	3.7	935	82	CNS006KX	AL068051 Drosophila
11	45.4	3.7	591	69	AM128786	AL188786 Drosophila
12	45.2	3.7	744	82	CNS0172K	AL108658 Drosophila
13	45.2	3.7	1101	83	CNS0175X	AL108460 Drosophila
14	45.2	3.7	315	51	AT172257	AL172257 605069D09
15	45.2	3.6	846	82	CNS010R3	AL059337 Drosophila
16	44.8	3.6	932	82	CNS0072Q	AL066742 Drosophila
17	44.4	3.6	843	82	CNS00C51	AL059666 Drosophila
18	43.8	3.5	427	51	AT173678	AL173678 sv32B04.Y
19	43.8	3.5	512	74	AV396794	AV396794 AV396794
20	43.8	3.5	521	74	AV397116	AV397116 AV397116
21	43.8	3.5	526	74	AV394827	AV394827 AV394827
22	43.6	3.5	1101	83	CNS016UE	AL107216 Drosophila
23	43.2	3.5	439	91	AQ084053	AQ084053 nbe0022M
24	43.2	3.5	798	82	CNS00A7A	AL055851 Drosophila
25	43.2	3.5	413	74	AV396947	AV396947 AV396947
26	41.8	3.4	414	80	AM286084	AM286084 LG1_261.B
27	41.8	3.4	465	80	AM287007	AM287007 LG1_261.B
28	41.8	3.4	530	74	AV396353	AV396353 AV396353
29	41.8	3.4	839	82	CNS004MB	AL054280 Drosophila
30	41.8	3.4	932	82	CNS0077Q	AL066742 Drosophila
31	41.6	3.4	506	74	AM180713	AM180713 Mga0863E
32	41.6	3.4	870	82	CNS006E2	AL064271 Drosophila
33	41.4	3.3	1201	82	CNS014B7	AL078875 Drosophila
34	41.4	3.3	639	60	AT1812194	AT1812194 605086P10
35	41.2	3.3	774	105	AQ327321	AQ327321 nbx0040D
36	41.2	3.3	910	82	CNS0060N	AL056652 Drosophila
37	41.2	3.3	605	46	AT161529	AT161529 486018G05
38	41.1	3.3	1101	82	CNS0175X	AL108460 Drosophila
39	40.8	3.3	384	64	AM054773	AM054773 ws59408.x
40	40.8	3.3	385	21	D48746	D48746 R1C51537A
41	40.8	3.3	437	74	AV395790	AV395790 AV395790
42	40.8	3.3	506	74	AV396405	AV396405 AV396405
43	40.6	3.3	574	91	AQ084186	AQ084186 LMAFV1.1
44	40.4	3.2	916	83	CNS015S0	AL105860 Drosophila
45	40.2	3.2	916	83	CNS015S0	AL105860 Drosophila

# ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	COMMENT
CNS0091P/c	925 bp	DNA	AL053013	1	GI:4934461	Drosophila melanogaster	03-JUN-1999	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		Determination of this BAC-end sequence was carried out as part of a

RESULT 2	LOCUS	DEFINITION	ACCESSION	VERSION	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	COMMENT
CNS0091P	925 bp	DNA	AL053013	1	GI:4934461	Drosophila melanogaster	03-JUN-1999	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		Determination of this BAC-end sequence was carried out as part of a

RESULT 3	LOCUS	DEFINITION	ACCESSION	VERSION	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	COMMENT
CNS0091P	925 bp	DNA	AL053013	1	GI:4934461	Drosophila melanogaster	03-JUN-1999	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		Determination of this BAC-end sequence was carried out as part of a

RESULT 4	LOCUS	DEFINITION	ACCESSION	VERSION	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	COMMENT
CNS0091P	925 bp	DNA	AL053013	1	GI:4934461	Drosophila melanogaster	03-JUN-1999	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		Determination of this BAC-end sequence was carried out as part of a

RESULT 5	LOCUS	DEFINITION	ACCESSION	VERSION	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	COMMENT
CNS0091P	925 bp	DNA	AL053013	1	GI:4934461	Drosophila melanogaster	03-JUN-1999	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		Determination of this BAC-end sequence was carried out as part of a

RESULT 6	LOCUS	DEFINITION	ACCESSION	VERSION	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	COMMENT
CNS0091P	925 bp	DNA	AL053013	1	GI:4934461	Drosophila melanogaster	03-JUN-1999	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutomo Oseguera and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source  
1..925  
Location/Qualifiers  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPc1-98"  
/clone="BAC19D16"  
/note="end : TET3"

BASE COUNT  
120 a 61 c 61 g 172 t 511 others

Query Match  
Best Local Similarity 11.9%; Pred. No. 0.055; Mismatches 142; Indels 0; Gaps 0;

QY	442	CCCCCTTGGCCCCGATGACGAGTCGCCGCTGACCTGCGCCAGGTTGGCCTTAC	501
DB	920	SCSCSCSBSBCSSSSSMTSSSBCSSSBSSTSSMSSSSSSSSSSSSSSSSSSSSS	861
QY	502	CACCTATGCGGACCAAGGGGGCGGCTGCTGATGGGGGGGGGGGGGGGGGGGGGG	561
DB	860	SACVKNMASSCGCCCGGACBCMCSSSSCCGSAKRVKFRASGAGKRGSGGASA	801
QY	562	CGCGCCGACCTGCTGTGATCGCGCGCCGACCGCGGCTTACAGCAGCCGATCCG	621
DB	800	SHSSSSACSSSSSSSSSACMSASSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS	741
QY	622	AACGGCATGGGGCGGACCTTACGCTTACGATCATCATCATCATCATCATCATCAT	681
DB	740	VSSASS	681
QY	682	GACGCCGAGTTCGCGCGCGGACGATCCACTCTACTCTACTCTACTCTACTCTACT	741
DB	680	SCSTASMSAARSS	621
QY	742	GCCGTAACAGCTGCGGACCTGATGTTGGGGCGGCTTGGTCCAGCGCCAGCCACC	801
DB	620	GGGGSVSSASSGMSVSSSSGGRSSGSGGGGGGSSSSSSSSSSSSSSSSSSSSSS	561
QY	802	A 802	
DB	560	R 560	

**AUTHORS** Genoscope.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
**COMMENT** - Web : www.genoscope.cns.fr  
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammose in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's pl and Est libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
**FEATURES**  
 source  
 1. 925  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="RPCI-98"  
 /clone="BAC19D16"  
 /note="end : TET3"  
**BASE COUNT** 120 a 61 c 61 g 172 t 511 others  
**ORIGIN**

Query Match 4.5%; Score 55.2; DB 82; Length 925;  
 Best Local Similarity 11.8%; Pred. No. 0.15;  
 Matches 41; Conservative 170; Mismatches 137; Indels 0; Gaps 0;

425 CCGCGAGCGGCGACTACCCCTGCTGGCCGATGAGCACTGCGGCTGCGACTGCGCG 484  
 Db SCGCGSSSCCBCCGCCSCSSSYCCSSBSBSCSTBSCGCCSKVCGTSCGSSSS 636  
 QY CCGCAGTTGGCGCTTACCACTGATGCAACCAAGGGGGCGCGGTGTGTGATGGCGG 544  
 Db CCGCGSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSS 696  
 QY 545 GGGTCCCCGGCTGCAACCGCGGACGTCGTGTGATGCGGCCCGGACCGCGGCTACA 604  
 Db 697 SSSSSSTSSSTSSSVSSGSSSTBSSGSSSSSTSSBBSCTSTSSSSSSSSSS 756  
 QY 605 ACGCAGCCCGCATCGCAAGCATGGGCGCGCATGCTTACGATCAACATCG 664  
 Db 757 TCSCCTCCCSYSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSS 816  
 QY 665 ACAACCTTGGCACTGCAACCGGAGTTCTGGCGGATCCACACTGCTACTGATCGG 724  
 Db 817 TYBMCYRSTSGCGSSSSGKGVTKCGCGCGSSSTNMBGTSSACSSSSSSSSSSSV 876  
 QY 725 CCTACGACTGAGGGTGGCGGCAACGTCGCGACCTGCGTATTTGGG 772  
 Db 877 SSSKSSASSSVSSGSSGVSNSSSASAKSSGSSGVS6GSGGSGSVS 924

**RESULT** 3  
**LOCUS** AV390505 535 bp mRNA EST 21-JAN-2000  
**DEFINITION** AV390505 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii  
 cDNA clone CM052B05\_r, mRNA sequence.  
**ACCESSION** AV390505  
**VERSION** AV390505.1 GI:6544721  
**KEYWORDS** EST.  
**SOURCE** Chlamydomonas reinhardtii.  
**ORGANISM** Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 Chlamydomonadaceae; Chlamydomonas.  
**REFERENCE** 1 (bases 1 to 535)

**AUTHORS** Asamizu, E., Nakamura, Y., Sato, S., Fukuzawa, H. and Tabata, S.  
**TITLE** A Large Scale Structural Analysis of cDNAs in a Unicellular Green Alga, Chlamydomonas reinhardtii. I. Generation of 3433 Non-redundant Expressed Sequence Tags  
**JOURNAL** DNA Res. 6, 369-373 (1999)  
**COMMENT** On Dec 20, 1995 this sequence version replaced gi:1135919.  
 Contact: Yasukazu Nakamura  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: ynakamu@kazusa.or.jp, URL: <http://www.kazusa.or.jp/gen/plant/>.  
**FEATURES**  
 source  
 1. 535  
 /organism="Chlamydomonas reinhardtii"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone="CM052B05\_r"  
 /clone\_lib="Chlamydomonas reinhardtii C9"  
 /dev\_stage="photoautotrophic growth"  
 /note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI"  
**BASE COUNT** 107 a 195 c 148 g 85 t 9  
**ORIGIN**

Query Match 4.2%; Score 51.8; DB 74; Length 535;  
 Best Local Similarity 51.5%; Pred. No. 0.62;  
 Matches 119; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

565 GCCGACGTCGTGTGATGCGCCCGGACCGCGCTTCACAGCGCCGCGATCGCCAC 624  
 Db 221 GTGACGAGGTGGCCATGAGACCTGCTCTTCCAGTACGCGGCTTACCGCGCTG 280  
 QY 625 GCGATGCGCGCAGCCGTTACGCTTACATCAACATGCAACTGCGGAACCTGAC 684  
 Db 281 CCGCTGGGCTGACCGCTGTGTGCGCCAGGCGCTTCAAGTACCGGAGCGGTGAC 340  
 QY 685 GCGAGTTCTGGCGGATGATGCACTGCTACTGATGCGCTGAGCTGAGAGGTGCC 744  
 Db 341 GCGGACTGGAACCGGCTGATGACACCTGCTTCAACGCGGCTTCAACATGTCGACG 400  
 QY 745 GTCAACGTGCGGACCTGATGTTGGCGGCTCTGTGTGCCAGGCGCCAC 795  
 Db 401 GCGGCTGTGGCTGCTTCACGCGGAGGCGCTCAAGAGATGACGCGCATG 451

**RESULT** 4  
**LOCUS** CNS0052P/C 844 bp DNA GSS 03-JUN-1999  
**DEFINITION** Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR11P16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
**ACCESSION** AL056652  
**VERSION** AL056652.1 GI:4932342  
**KEYWORDS** GSS.  
**SOURCE** fruit fly.  
**ORGANISM** Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
**REFERENCE** Genoscope.  
**AUTHORS** Direct Submission  
**TITLE** Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
**JOURNAL** - Web : www.genoscope.cns.fr  
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammose in Pieter de Jong's laboratory in the Department of

NY. The library is named RPci-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

BASE COUNT	ORIGIN
261 a	112 c 92 g 35 t 344 others

BASE COUNT	261 a	112 c	92 g	35 t	344 others
ORIGIN					

## ORIGIN

Query Match	4.1%	Score 50.2	DB 82	Length 844
Best Local Similarity	17.2%	Pred No. 1.3		
Matches 65	Conservative 161	Mismatches 150	Indels 1	Gaps 1

QY 257 CCGAGCGATTATTTATGCTCAGGTCAGCAAGAACCAGTATGCGCGGAATATACGCCGCTGC 312

Db 768 CCMVCCCTTTHTTTCMTAAMRMTAMRMMAMMMATYKCYMMAMVAAMVYTAAMAMRM 705

QY 317 GACACGGGCGACAGATCTTGTTCACGTTCTTGCAATTTGGCCGCGCTCAGCGCTTGACACGATG 378

Db 708 NNVCYSBSCNBTHTTYTCNTCVSSBSSBTTCTWHTNSNTYMTVYVTTHTHTNTC 644

QY 377 CCGTGTGGATTCGGGCAACACGTCATGCTCTACGAGACCGCTCAGACCGCGGACGGC 437

Db 648 CYYCCCCCYCSCCBYBCNTCSYYSYBSCGSSSSSSSSCGSSSBYSOCCBCCBGY 583

QY 437 CACTAC -CCCTGTTGCCCCGATAGCAAGTGCAGCGGTCACTCGGCGGCCACAGTTGGC 492

Db 588 CGCTSBVCSBSSSTBSBVSYSGBSSTSBTGCKBSBSSBSCBSBSCCTCYSTYCY 522

QY 496 GCTTACCACCTGATGCGAACCCAAAGGGGCGCCGAGTGCTGATGGCGGGGTGCCCGC 552

Db 528 SSCBCYCBSCBMTBSYBSBSGSGBTGBSBBSGKSGSGKSGGGBKSGGKGYCYBITYC 462

QY 556 GTCCGACCGGCCGACGTCGTGTGATCGGCGCCGCGCACCCCGGCTCAACAGCACCGCCG 613

Db 468 CSBSBSCBSSSSSTSKCCSGSSSTKCGSSBBSBGGSSSKTSGSSSCBBSGSSSTGSSBSS 409

QY 616 ATGCGCAACGCGATGGG 632

Db 408 KKISTYSTSSSSGCGTG 392

RESULT	5			
CNS0060N/c				
LOCUS	CNS0060N	910 bp	DNA	GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BACR47471 of RPCT-98 library from Drosophila melanogaster (fruitfly), genomic survey sequence.			

ACCESSION	AL065629	
VERSION	AL065629.1	GI:49446988
KEYWORDS	GSS.	
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster	

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT Determination of this BAC-end sequence was carried out as part of a

collaboration with the BerkeleyDrosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org>. The BDGP Drosophila melanogaster BAC library was prepared by Kazutuyo Oosawa and Aaron Mammoset in Pieter de Jong's Laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <http://bacpac.med.buffalo.edu/drosophila/bac.htm>.

## FEATURES

	source	I . 910			
	/organism	"Drosophila melanogaster"			
	/db_xref	"taxon:7227"			
	/clone-1b	"RPCI-98"			
	/clone-	"BACR14J21"			
	/note-end	: 17"			
BASE COUNT		202 a	63 c	112 g	198 t
ORIGIN					335 other

BASE COUNT	202 a	63 c	112 g	198 t	335 others
ORIGIN					

ORIGIN

```

Query Match      4.08; Score 49; DB 82; Length 910;
Best Local Similarity 20.58; Pred. NO. 2.3;
Matches 47; Conservative 97; Mismatches 85; Indels 0; Gaps

```

DQ  
DB

713 SRVMAAGSASASSSSSSASSASSASSA\$SA\$MVSICRA\$VA\$AA\$AAA\$KRS\$RASA 65\$

DQ 453 CCGCATGACGAATGCCGGGTGCATCGCCGCCCAAGTTGG...CTTACCATTGATGCG 512\$

DB 653 YMAVAAAS\$SA\$VASSG...VSCSSPCCGCSASS\$SA\$ASG...SSSSSSSSCGCGSAS 594\$

DQ 513 AACCCAAGGGCCCCCGGTGTCTCATTATGGCGGGGGTGTCGCCGCGTGAACCGGGCGACGT 572\$

DB 593 CA\$SSSRSGSCCGGSYVCSSSCGSSSSSSSSSSSSSGCSSSCCCSC\$CACRCCCC 534\$

DQ 573 CGAGTGTATCGCGCGCGACCGCGGCGCTACAACGCAAGCCCGCATGCG 621\$

DB 533 CRCSGCGCCCMSSSCCSCCTCTA\$MTCAMAVAAHACCCSCCCCDC 485\$

RESULT	6
CNS010R3/c	
LOCUS	CNS010R3 845 bp DNA GSS
DEFINITION	Drosophila melanogaster genome survey sequence 17 end of BAC
ACCESSION	BAC04N13 of DrosBAC library from Drosophila melanogaster (fruitfly), genomic survey sequence.
	AI099337

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE
AL099337	AL099337.1	GI:5610948	GSS.	fruit fly, <i>Drosophila melanogaster</i>	
				<i>Eukaryota; Metazoa; An</i>	
				<i>pterygota; Neoptera; E</i>	
				<i>muscomorpha; Ephydroid</i>	
				<i>1 (bases 1 to 846)</i>	

REFERENCE AUTHORS TITLE JOURNAL	COMMENT
3 (bases 1 to 845) Genoscope. Direct Submission Submitted (23/JUL-1999) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail: seqtel@genoscope.cns.fr Web : www.genoscope.cns.fr)	Determination of this BAC-end sequence was carried out as part of a

COMMENT  
Determination of this BAC-end sequence was carried out as part of collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk/>. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CEPH (Centre d'Etude du Polymorphisme humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector



FEATURES	pbeOBAC11.	Location/Qualifiers	
source	1. 846	/organism="Drosophila melanogaster"	
		/plasmid="pbeOBAC11"	
		/db_xref="taxon:7227"	
		/clone_lib="DrosBAC"	
		/clone="BACN04N13"	
		/note="end : T7"	
BASE COUNT	300 a 47 c 40 g 175 t 284 others		
ORIGIN			
	Query Match	3.8%; Score 46.6; DB 82; Length 846;	
	Best Local Similarity 13.0%; Pred. No.6.6;		
Matches	33; Conservative 127; Mismatches 93; Indels 0; Gaps 0		
QY	363 TCGTTCACCCGATGCGCTGTTCGATTCGCCGACACGTCATTGCTACGACACGTCGA	422	
DB	842 TCTCTSSSTSSCGSSSBSTBTSTTSTSSASSSSSSTSSSTSSSTSSSTSBTBT	783	
QY	423 GACCCGCGACGCGGCACCTCCCTGCTGCCGATAG/GAAGTCGCGGTGACACGCG	482	
DB	782 SBSTSTSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSST	723	
QY	483 CGCCACGTTGGCGCTTACCACTGATGCAACCCAAAGGGCGCGGTGCTGATGG	542	
DB	722 SBSTGCTSTSBTBTBTBTSBSSBSSSSSTSTBTBSSSSSBSTSTSTSS	663	
QY	543 CGGGGTGCGCGCGGTGAACCGCGCGCGGTGATGCGGGCGCGACCGCGGCTA	602	
DB	662 SSSBTSSTSSSBSTSTTTTSTTSSCGSSSSSTSSSTSSSSSBSTSSCGSSSSCCT	603	
QY	603 CAACGACGCGCGC 615		
DB	602 CGCCSCGCCGCC 590		
RESULT 7			
LOCUS	CNS0181E	1101 bp	DNA
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC		
	BACN37H05 of DrosBAC library from Drosophila melanogaster (fruit		
	fly), genomic survey sequence.		
ACCESSION	AL108764		
VERSION	AL108764.1	GI:5629068	
KEYWORDS	GSS.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1101)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage		
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
	- Web : www.genoscope.cns.fr)		
COMMENT	Determination of this BAC-end sequence was carried out as part of		
	collaboration with the European Drosophila genome Project (EDGP) -		
	http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC		
	library (Dros BAC) was made by Alain Billaud at CEPH (Centre		
	d'Etude du Polymorphisme Humain) with funding provided by a MRC		
	project grant. The DNA was prepared from embryos by Alain Bucheton		
	and Genevieve Payan. It has been constructed in the vector		
	pbeOBAC11.		
FEATURES			
source	Location/Qualifiers		
	1. 1101		
	/organism="Drosophila melanogaster"		
	/plasmid="pbeOBAC11"		
	/db_xref="taxon:7227"		
	/clone_lib="DrosBAC"		
	/clone="BACN37H05"		

BASE COUNT	61 a	215 c	301 g	235 t	289 others
ORIGIN	/note="end : SP6"				
Query Match	3.8%: Score 46.6; DB 83; Length 1101;				
Best Local Similarity	40.2%: Pred. No. 6.9;				
Matches 101; Conservative	40; Mismatches 109; Indels 1; Gaps 1				
QY	389	CGGACACACGTCATTTGGCTACGAGACCGTTCACAGACCCGCCACGCGGACGACTACCCCTCC	448		
Db	308	CCCCCCTVVVVVGMMAASCCCAVASSSCCCSCCCSCCCSCCCGCGCGSCACSCSCSCSG	249		
QY	449	TTGCCCCGATGAGCGAAG:CGCGGTGCAGCTCGCGCCGCCAGGTGTGGCGGTTACCACTGA	508		
Db	248	AAGMGSMGAGVGCACAGAGGCGCGAGACCCGSCACSCCCSSCCGSSCCSSSSS	189		
QY	509	TGCGAACCCAGAGGGGCGCGGTGTGCTGATGGCGGGGTGCCCGGTGCAACCGCCG	568		
Db	188	CGCGACCCSSSAGGGGCGCCSGCGGCCACCCCGCCSGAGCTVAGCCGCCAC-CCMYGGGSG	130		
QY	569	AGCTGTGATGTATCGCGCGCGCCGCGCGGTTCACAGACCCGCCGATCGCAACGGCA	628		
Db	129	ACGCAAGGAGGAGAGAGGCGCCGAGCCCGCCGACCCCGCCGCGCACGCGCGGGA	70		
QY	629	TGGCGCGCAAC	639		
Db	69	AGCGCCCAAC	59		
RESULT	8				
CNS006XK					
LOCUS	CNS006XK	935 bp	DNA	GSS	03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC #				
	BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit				
	fly), genomic survey sequence.				
ACCESSION	AI066051				
VERSION	AI066051.1 GI:4945019				
KEYWORDS	GSS.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;				
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 935)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr				
COMMENT	- Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .				
FEATURES	Location/Qualifiers				
Source	1..935 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPCI-98" /clone="BACR14N09" /note="end : T7"				
BASE COUNT	257 a	170 c	162 g	96 t	250 others







/note="Organ: Kernel; Vector: PAD-GAL4-2"; Site\_1: EcorI;  
Site\_2: XhoI; Kernel endosperm cDNA library from Schmidt,  
lab"

BASE COUNT 61 a 105 c 106 g 42 t 1 others

Query Match 3.6%; Score 45; DB 51; Length 315;  
Best Local Similarity 50.7%; Pred. No. 12;  
Matches 108; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 391 GGCACACGCGATGCTGCTGACGACGCGCGACGCGCGCTGCTT 450  
DB 58 GGCATGCGCTCTCATCCAAAGACACAGCTGCGTGGCCAGAACTTTAAACACCCC 117  
QY 451 GCGCCGATGAGCAGAGTCCGCGCTGCGCTGCGCGCGCGCTTACACCTGATG 510  
DB 118 GCGCTTAAAGCAGCAGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 177  
QY 511 CGAACCAGAGGCGCGCGCTGCTGATGCGCGCGCGCGCGCGCGCGCGCGCGCG 570  
DB 178 CGTGGCGACAGGCGCGCGCGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 237  
QY 571 GTCGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 603  
DB 238 GTCGGCGGTGCG 270

RESULT 15

CNS010RJ

LOCUS 846 bp DNA GSS 26-JUL-1999  
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC  
BACN04N13 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION AL099337.1 GI:5610948

VERSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster  
fruit fly.  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr  
- Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billand at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos of Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.

FEATURES  
source

1. 846  
Location/Qualifiers  
/organism="Drosophila melanogaster"  
/plasmid="pBelobAC11"  
/db\_xref="taxon:7227"  
/clone\_lib="DrosBAC"  
/clone="BACN04N13"  
/note="end : 77"  
BASE COUNT 300 a 47 c 40 g 175 t 284 others  
ORIGIN

Query Match 3.6%; Score 45; DB 82; Length 846;  
Best Local Similarity 16.2%; Pred. No. 14;  
Matches 38; Conservative 109; Mismatches 88; Indels 0; Gaps 0;

QY 506 TGATGGAGACCCAGGCGCGGTGTGTGTATGGCJGGGTGCCCGCGGTGGAACGG 565

DB 574 KGACCGRARRSSAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 633  
QY 566 CCACGCTGCTGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 625  
DB 634 SSGSSSSAAASAAAGASSSVSSASSSVSSSSSSAAASSSASASVSSSVSSAAVSSASS 693  
QY 626 GCATGGCGCGACCGCTTACGCTTACAGATCAGACATGACAACTTGGCAACTCGAGC 685  
DB 694 SSYSVSSSSAAVAAVAVVSSASASVSSSVSSGCAASASASVSTSSSSAASSSSSSG 753  
QY 686 CCGAGTTTGGCGCGCGCGATCCACTCGCTACTCATCGGCTACGACTCGAGGG 740  
DB 754 SSASGASSSSAAASAAASSSAAASVAVAVSSASSSVSSASSSSSSSSSS 808

Search completed: June 22, 2000, 12:07:29  
Job time: 10363 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:52:44 ; Search time 8627.09 Seconds

(without alignments)  
-138.469 Million cell updates/sec

Title: US-09-362-485-5  
1228  
Sequence: 1 ATCTTGAGATTAATCGAAC.....GGAGTAAGGAGAGATGA 1228

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenDbml: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_cm: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pl1: \*  
8: gb\_pl2: \*  
9: gb\_pl3: \*  
10: gb\_pr1: \*  
11: gb\_pr2: \*  
12: gb\_pr3: \*  
13: gb\_ro: \*  
14: gb\_sts: \*  
15: gb\_sy: \*  
16: gb\_un: \*  
17: gb\_vl: \*  
18: em\_fun: \*  
19: em\_hum1: \*  
20: em\_hum2: \*  
21: em\_in: \*  
22: em\_om: \*  
23: em\_or: \*  
24: em\_ov: \*  
25: em\_pat: \*  
26: em\_ph: \*  
27: em\_pl: \*  
28: em\_ro: \*  
29: em\_sts: \*  
30: em\_sy: \*  
31: em\_un: \*  
32: em\_vl: \*  
33: gb\_htg1: \*  
34: gb\_htg2: \*  
35: gb\_in1: \*  
36: gb\_in2: \*  
37: em\_ba1: \*  
38: em\_ba2: \*  
39: em\_hum3: \*  
40: gb\_pr4: \*  
41: gb\_htg3: \*  
42: gb\_htg4: \*  
43: gb\_htg5: \*  
44: gb\_htg6: \*

45: gb\_htg7: \*  
46: em\_htg1: \*  
47: em\_htg2: \*  
48: em\_htg3: \*  
49: em\_hum5: \*  
50: gb\_pl3: \*  
51: gb\_pr5: \*  
52: gb\_htg8: \*  
53: gb\_htg9: \*  
54: gb\_htg10: \*  
55: gb\_htg11: \*  
56: gb\_htg12: \*  
57: gb\_htg13: \*  
58: gb\_htg14: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1228	100.0	1228	5	A87608 Sequence 5
2	1228	100.0	1328	5	A89748 Sequence 5
3	1228	100.0	1237	5	A87607 Sequence 4
4	1228	100.0	1237	5	A89747 Sequence 4
5	1228	100.0	56414	1	MTV002
6	1226.4	99.9	1236	5	A87613 Sequence 10
7	1226.4	99.9	1236	5	A89753 Sequence 10
8	1224.4	99.7	1235	5	A87606 Sequence 3
9	1224.4	99.7	1235	5	A89745 Sequence 3
10	1213.4	98.8	1235	5	A87609 Sequence 6
11	1213.4	98.8	1235	5	A89749 Sequence 6
12	1212.8	98.8	1235	5	A87611 Sequence 8
13	1212.8	98.8	1235	5	A89751 Sequence 8
14	1211.8	98.7	1229	5	A87610 Sequence 7
15	1211.8	98.7	1229	5	A89750 Sequence 7
16	1210.4	98.6	1245	5	A87605 Sequence 2
17	1210.4	98.6	1245	5	A89745 Sequence 2
18	1210.4	98.6	1260	5	A87604 Sequence 1
19	1210.4	98.6	1260	5	A89744 Sequence 1
20	1210.4	98.6	2584	1	MTALADH
21	1206.4	98.2	1209	5	A87612 Sequence 9
22	1199.4	97.7	1208	5	A89752 Sequence 9
23	1192.4	97.1	1194	2	MT092472
24	538.4	43.8	40745	1	SC151
25	538.4	43.8	40745	1	SC151
26	467.6	38.1	682	5	A89758 Sequence 15
27	467.6	38.1	682	5	A89752 Sequence 15
28	377	30.7	8046	2	AF049107 Myxococcus
29	351.2	28.6	11514	2	AE002028 Deinococcus
30	331.4	27.0	2780	1	RLE238118 Rhizobium
31	330.8	26.9	1585	2	AF070715 Vibrio pr
32	312.8	25.5	1736	1	BACALDHC
33	293	23.9	1477	2	AF070715 Shewanella
34	288.4	23.5	2113	1	D37807 Phormidium
35	277.2	22.6	1125	5	E01847 DNA encodin
36	258	21.0	137740	1	D90900 Synchocyst
37	243.2	19.8	1375	1	AB013821 Enterobac
38	243	19.8	1776	1	BACALDH
39	235.6	19.2	1709	1	BACALDH
40	235.6	19.2	12316	1	B5282015
41	235.6	19.2	217420	1	B5282015
42	218.2	17.8	1714	2	AF070714 Bacillus su
43	107.2	8.7	10632	2	AE001565 Helicobac
44	99.2	8.1	10512	2	AE000639 Helicobac
45	74.2	6.0	3160	1	RU05294 Rhodospirillum

#### ALIGNMENTS





No.	Score	Match	Length	DB	ID	Description
1	57.4	4.6	925	82	CNS0091P	AL053013 Drosophila
2	55.2	4.2	925	82	CNS0091P	AL053013 Drosophila
3	51.8	4.0	535	74	AV390505	AV390505 AV390505
4	49.0	3.8	844	82	CNS0056N	AL056625 Drosophila
5	46.6	3.8	844	82	CNS0056N	AL056625 Drosophila
6	46.6	3.8	844	82	CNS0056N	AL056625 Drosophila
7	46.6	3.8	844	82	CNS0056N	AL056625 Drosophila
8	46.4	3.8	844	82	CNS0056N	AL056625 Drosophila
9	45.8	3.7	935	82	CNS0065X	AL066051 Drosophila
10	45.4	3.7	591	65	AM128786	AM128786 es7405.y
11	45.4	3.7	744	83	CNS0172K	AL108460 Drosophila
12	45.2	3.7	1101	83	CNS0175Y	AL108460 Drosophila
13	45.2	3.6	315	51	AV1712257	AL101589 Drosophila
14	45.2	3.6	846	82	CNS0108J	AL099337 Drosophila
15	44.8	3.6	843	82	CNS00720	AL066742 Drosophila
16	44.8	3.6	843	82	CNS00720	AL066742 Drosophila
17	43.8	3.5	427	51	AT736678	AL056666 Drosophila
18	43.8	3.5	512	74	AV396794	AV396794 AV396794
19	43.8	3.5	521	74	AV397116	AV397116 AV397116
20	43.8	3.5	526	74	AV394827	AV394827 AV394827
21	43.8	3.5	645	82	CNS01213	AL101589 Drosophila
22	43.6	3.5	1101	83	CNS016DE	AL107216 Drosophila
23	43.2	3.5	798	82	CNS0040J	AL058851 Drosophila
24	43.2	3.5	798	82	CNS0040J	AL058851 Drosophila
25	43.2	3.5	798	82	CNS0040J	AL058851 Drosophila
26	41.8	3.4	413	74	AV396947	AV396947 AV396947
27	41.8	3.4	465	80	AM287007	AM287007 L61_264.C
28	41.8	3.4	530	74	AV396353	AV396353 AV396353
29	41.8	3.4	839	82	CNS004NB	AL054280 Drosophila
30	41.8	3.4	932	82	CNS00720	AL066742 Drosophila
31	41.8	3.4	1101	83	CNS0175Y	AL108460 Drosophila
32	41.6	3.4	506	74	AM180713	AM180713 M9AC863F
33	41.6	3.4	870	82	CNS006E2	AL064271 Drosophila
34	41.4	3.4	1101	82	CNS0014J	AL078875 Drosophila
35	41.4	3.4	1201	83	CNS014B7	AL103945 Drosophila
36	41.2	3.3	639	60	AO3227321	AO3227321 nbx0040D
37	41.2	3.3	774	105	AO3227321	AO3227321 nbx0040D
38	41.2	3.3	910	82	CNS0060N	AL056625 Drosophila
39	41.2	3.3	910	82	CNS0060N	AL056625 Drosophila
40	40.8	3.3	384	64	AM054773	AM054773 466018G05
41	40.8	3.3	385	21	D48746	AM054773 466018G05
42	40.8	3.3	437	74	AV395790	AV395790 AV395790
43	40.6	3.3	506	74	AV396405	AV396405 AV396405
44	40.6	3.3	574	91	AO848186	AO848186 LMAJFV1_1
45	40.2	3.3	916	83	CNS0155Q	AL105860 Drosophila

ALIGNMENTS

RESULT 1  
LOCUS CNS0091P/c  
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR19D16 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL053013.1 GI:4934461  
VERSION AL053013  
KEYWORDS fruit fly.  
SOURCE Drosophila melanogaster  
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 925)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source  
location/Qualifiers  
1..925  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPI-98"  
/clone="BACR19D16"  
/note="end : TET3"

BASE COUNT 120 a 61 c 61 g 172 t 511 others  
ORIGIN

Query Match 4.6%; Score 57.4; DB 82; Length 925;  
Best Local Similarity 11.9%; Pred. No. 0.06;  
Matches 43; Conservative 176; Mismatches 142; Indels 0; Gaps 0;

QY 441 CCCTGCTTCCCGATGACCAATGCGCGGTCGACGCGCCAGGTTGCGCTTAC 500  
DB 920 SCSCSCSBCSSSMTSTSSBSSBSSSTSSMSBSSSSSSSSSSSSSSSS 861  
QY 501 CACCTGATGCGAACCAGGCGCGGTCGATGCGCGGTCGCGCGGTCGCGA 560  
DB 860 SACKKMASSCCCGCGAABCCSSSSCCSASARVKNASGAKRGGSAGSA 801  
QY 561 CCGGCGACGCTGCTGATGCGCGCGGCGCGGTCGATGCGCGGTCGCGG 620  
DB 800 SHSSSSACBSSSSSSSCASGWSASSSSSSSSSSSSSSSSSSSSSSSS 741  
QY 621 AAGGCGATGCGCGCGGCGGTCGATGCGGTCGATGCGGTCGCGGTCGCG 680  
DB 740 VSSASASSSSSSSVSCSVSSBSSSSSSSSSSSSSSSSSSSSSSSSSS 681  
QY 681 GACCGCGATGCTGCGCGGCGGTCGACACTGCTACTGCGGTCGCGG 740  
DB 680 SCSTASMSAARSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 621  
QY 741 GCGCTCAAGCTCCGACCTGATGCGCGGCGGTCGCTGCTGCGCGGCG 800  
DB 620 GSGSVASSGMSVSSGSGSGGCGGCGGCGGCGGCGGCGGCGGCGGCG 561  
QY 801 A 801  
DB 560 R 560

RESULT 2  
LOCUS CNS0091P  
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR19D16 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL053013.1 GI:4934461  
VERSION AL053013  
KEYWORDS fruit fly.  
SOURCE Drosophila melanogaster  
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 925)





```

FEATURES
  source
    pBelobAC11.
      Location/Qualifiers
        1..846
          /organism="Drosophila melanogaster"
          /plasmid="pBelobAC11"
          /db_xref="taxon:7227"
          /clone_lib="DrosBAC"
          /clone="BACN04N13"
          /note="end : 17"

BASE COUNT      300 a      47 c      40 g      175 t      284 others

ORIGIN
Query Match
Best Local Similarity 13.0%; Score 46.6; DB 82; Length 846;
Matches 33; Conservative 127; Mismatches 93; Indels 0; Gaps 0;

QY 362 TCGTTGACCGATGCGTGTGATTCGGACACGCTCAATGCTACGACGCTCCA 421
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 842 TTCTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSS 783

QY 422 GACCGCGACGCGACGCTGCTGCTGCGCGATGAGCGAATGCGCGGCTGCTGCG 421
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 782 SBSTSTSTSTSSSTSTSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSS 723

QY 482 CGCCAGGTTGGCGCTTACCATGATGACCAAGGCGCGCGGTGCTGATGG 541
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 722 SBSTSTSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSS 663

QY 542 CGGGGTGCGCGGCTGACACGCGCGCGCTGCTGATGCGCGCGCGCTA 601
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 662 SSSSBSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSS 603

QY 602 CAACGACGCGCG 614
    ||| : : : : :
Db 602 CGCCSCGCGCGCG 590

RESULT 7
LOCUS CNS0181E/c 26-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC #
            BACN37H05 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION AL108764.1 GI:5629068
VERSION AL108764.1
KEYWORDS GSS.
SOURCE Drosophila melanogaster
  ORGANISM fruit fly.
    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
    Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
    Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 1101)
REFERENCE
  Genoscope.
  Submitted (23-JUN-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  Web : www.genoscope.cns.fr)
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the European Drosophila Genome Project (EDGP) -
  http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
  library (Dros BAC) was made by Alain Billaud at CEPH (Centre
  d'Etude du Polymorphisme Humain) with funding provided by a MRC
  project grant. The DNA was prepared from embryos provided by Alain Bucheton
  and Genevieve Payan. It has been constructed in the vector
  pBelobAC11.
FEATURES
  source
    Location/Qualifiers
      1..1101
        /organism="Drosophila melanogaster"
        /plasmid="pBelobAC11"
        /db_xref="taxon:7227"
        /clone_lib="DrosBAC"
        /clone="BACN37H05"

```

```

BASE COUNT      61 a      215 c      301 g      235 t      289 others

ORIGIN
Query Match
Best Local Similarity 40.2%; Score 46.6; DB 83; Length 1101;
Matches 101; Conservative 40; Mismatches 109; Indels 1; Gaps 1;

QY 388 CGGCACACGATGCTGCTGACGACCGCTCAGACCGCGCGCGCATACCTTCG 447
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 308 CCCCMCVVVVMMVAASCCCAVASCSCCSCSCSCSCSCSCSCSCSCSCSC 249

QY 448 TTGCCCCGATGAGCGAAGTCCGCTGACGCGCGCGCGCGCGCGTTCGCTGAC 507
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 248 AGMCGSMGAGGACAGGCGCGAGACCGCGCGCGCGCGCGCGCGCGCGCG 189

QY 508 TCGAACCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 567
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188 CGGACCGCCSSAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 130

QY 568 AGTGTGTGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 627
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 129 AGCMAAGGAGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 70

QY 628 TGGCGCGGACG 638
    ||| : : : : :
Db 69 AGGCGCGGACG 59

RESULT 8
LOCUS CNS006XK 935 bp DNA 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION AL066051.1 GI:4945019
VERSION AL066051.1
KEYWORDS GSS.
SOURCE Drosophila melanogaster
  ORGANISM fruit fly.
    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
    Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
    Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 935)
REFERENCE
  Genoscope.
  Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  Web : www.genoscope.cns.fr)
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the Berkeley Drosophila Genome Project (BDGP).
  The BDGP is constructing a physical map of the Drosophila
  melanogaster genome using these BACs. For further information
  please see http://www.fruitfly.org The BDGP Drosophila
  melanogaster BAC library was prepared by Kazuhiro Oseguwa and
  Aaron Mammoser in Pieter de Jong's laboratory in the Department of
  Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
  NY. The library is named RPCI-98 and was constructed by partial
  EcoRI digestion of Drosophila DNA provided by the BDGP from the
  isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
  P1 and EST libraries. A more detailed description of the library
  and how to order individual BAC clones, the entire library, or
  filters for hybridization from the BACPAC Resource Center can be
  found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
  source
    Location/Qualifiers
      1..935
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone_lib="RPCI-98"
        /clone="BACR14N09"
        /note="end : 17"

BASE COUNT      257 a      170 c      162 g      96 t      250 others

```



/db\_xref="taxon:7955"  
/clone\_lib="zebrafish Washu MPMG EST"  
/sex="mixed"  
/tissue\_type="26 somite embryos, adult livers, shield  
stage embryos"  
/lab\_host="XLI-blue MRF"  
/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; 1st  
strand cDNA was primed with a Not I - oligo(dT)15 primer  
(5'-GACAGTTCAGTTCAGCGAGCGCCGCTTTTCTTTT3');  
double-stranded cDNA was ligated to Sal I adaptors (BR),  
digested with Not I and cloned into the Not I and Sal I  
sites of the pSPORT1 vector (BR). Library was constructed  
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck  
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST  
analysis were selected following oligonucleotide  
hybridization fingerprinting of arrayed clones from  
zebrafish late somitogenesis (26 ss), adult liver or  
embryonic shield stage (5.6 h) libraries. Fingerprint  
data were used to computationally cluster cDNAs, and a  
single cDNA from each cluster was chosen for sequencing.  
In some cases multiple members of the same cluster were  
sequenced to assess clustering parameters or single clones  
were sequenced additional times to assess quality  
control."

BASE COUNT 119 a 169 c 156 g 147 t  
ORIGIN

Query Match 3.7%; Score 45.4; DB 69; Length 591;  
Best Local Similarity 50.2%; Pred. No. 11;  
Matches 112; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

OY 2 TCTTCAGATTATGAACTTCTTCACATGAGCGTACATATGAGAGGGGTATCA 61  
DB 185 TCTTAGAGCATATCAAGCCCTCAACCGCTCAGTACGAGCATCTCTTCAAAACAC 244  
OY 62 TCGCGTGTGATTCGACGAGACCAAAACACGAAATTCGGGTGCCATCACCCGG 121  
DB 245 TCACCGTCGCGCTCTTCAAGAAATCTTCAAGATGAGCGTGCAGTGCATCTCTCTG 304  
OY 122 CCGGGTGCAGACTAACCCCTCGTGGCCATGAGTGTCTATCCAGGAGGTGCCGGAG 181  
DB 305 CCGGAGTGAACCTCTCATCAAGACAGGGCTTAAATGTGGTGGTCCGGCGGGAG 364  
OY 182 AGGCGTCCGCTATCACGACGCGGATTTCAAGCGGACGGCC 224  
DB 365 ATCTGCCAAGTCTCTGATGATGATGATACCAAAAGCAGGAGC 407

RESULT 11  
CNS0172K 744 bp DNA GSS 26-JUL-1999  
LOCUS  
DEFINITION Drosophila melanogaster genome survey sequence Sp6 end of BAC  
BACN37B10 of Drosophila library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL108698  
VERSION AL108698.1 GI:5629002  
KEYWORDS  
SOURCE GSS.  
ORGANISM fruit fly.  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 (bases 1 to 744)  
Genoscope.  
AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seq@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre

d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelosBAC11.

FEATURES  
source Location/Qualifiers  
1..744  
/organism="Drosophila melanogaster"  
/plasmid="pBelosBAC11"  
/db\_xref="taxon:7227"  
/clone\_lib="DrosBAC"  
/clone="BACN37B10"  
/note="end : Sp6"

BASE COUNT 20 a 191 c 245 g 105 t 183 others  
ORIGIN

Query Match 3.7%; Score 45.4; DB 83; Length 744;  
Best Local Similarity 33.8%; Pred. No. 12;  
Matches 92; Conservative 42; Mismatches 138; Indels 0; Gaps 0;

OY 312 CCTGGACAGCGGCGCATCTGTCACGTTCTGCAATTTGCCGCTCAGTCTTCAC 371  
DB 175 CCCCTGAGATCTCTCTCTTGTGCTGCTGCGCTGCGCTGCTGCTGCTGCTGCTG 234  
OY 372 GATGCGTTGTTGATTCGCGACACGATGCTACGAGACGCTCAGACCGCGAC 431  
DB 235 SYGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 294  
OY 432 GCGGACTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 491  
DB 295 SCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 354  
OY 492 GCGCTTACCATGATGAGACCAAGGCGCGCGCTGCTGCTGCTGCTGCTGCTG 551  
DB 355 NNCCNCCCGCCG 414  
OY 552 GCGCTGCAACCG 583  
DB 415 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 446

RESULT 12  
CNS0175Y 1101 bp DNA GSS 26-JUL-1999  
LOCUS  
DEFINITION Drosophila melanogaster genome survey sequence Sp6 end of BAC  
BACN37L08 of Drosophila library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL108460  
VERSION AL108460.1 GI:5628764  
KEYWORDS  
SOURCE GSS.  
ORGANISM fruit fly.  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 (bases 1 to 1101)  
Genoscope.  
AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seq@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelosBAC11.

FEATURES  
source Location/Qualifiers  
1..1101  
/organism="Drosophila melanogaster"  
/plasmid="pBelosBAC11"







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:54:29 ; Search time 8627.09 Seconds

(without alignments)  
-138.582 Million cell updates/sec

Title: US-09-362-485-7

Perfect score: 1229  
Sequence: 1 ATCTTGCAATTAATGAC.....GAGTAGGACGATGATG 1229

Scoring table: IDENTITY\_NDC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_bal: \*  
2: gb\_bal: \*  
3: gb\_bal: \*  
4: gb\_bal: \*  
5: gb\_bal: \*  
6: gb\_bal: \*  
7: gb\_bal: \*  
8: gb\_bal: \*  
9: gb\_bal: \*  
10: gb\_bal: \*  
11: gb\_bal: \*  
12: gb\_bal: \*  
13: gb\_bal: \*  
14: gb\_bal: \*  
15: gb\_bal: \*  
16: gb\_bal: \*  
17: gb\_bal: \*  
18: gb\_bal: \*  
19: gb\_bal: \*  
20: gb\_bal: \*  
21: gb\_bal: \*  
22: gb\_bal: \*  
23: gb\_bal: \*  
24: gb\_bal: \*  
25: gb\_bal: \*  
26: gb\_bal: \*  
27: gb\_bal: \*  
28: gb\_bal: \*  
29: gb\_bal: \*  
30: gb\_bal: \*  
31: gb\_bal: \*  
32: gb\_bal: \*  
33: gb\_bal: \*  
34: gb\_bal: \*  
35: gb\_bal: \*  
36: gb\_bal: \*  
37: gb\_bal: \*  
38: gb\_bal: \*  
39: gb\_bal: \*  
40: gb\_bal: \*  
41: gb\_bal: \*  
42: gb\_bal: \*  
43: gb\_bal: \*  
44: gb\_bal: \*

45: gb\_bal: \*  
46: gb\_bal: \*  
47: gb\_bal: \*  
48: gb\_bal: \*  
49: gb\_bal: \*  
50: gb\_bal: \*  
51: gb\_bal: \*  
52: gb\_bal: \*  
53: gb\_bal: \*  
54: gb\_bal: \*  
55: gb\_bal: \*  
56: gb\_bal: \*  
57: gb\_bal: \*  
58: gb\_bal: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1228	99.9	1229	5	A87610 Sequence 7
2	1228	99.9	1229	5	A87610 Sequence 7
3	1228	99.9	1235	5	A87611 Sequence 8
4	1228	99.9	1235	5	A87611 Sequence 8
5	1235.4	99.7	1235	5	A87609 Sequence 6
6	1235.4	99.7	1235	5	A87609 Sequence 6
7	1235.4	98.9	1236	5	A87613 Sequence 10
8	1235.4	98.9	1236	5	A87613 Sequence 10
9	1235.8	98.8	1237	5	A87607 Sequence 4
10	1235.8	98.8	1237	5	A87607 Sequence 4
11	1235.8	98.8	1237	5	A87607 Sequence 4
12	1235.4	98.7	1235	5	A87606 Sequence 3
13	1235.4	98.7	1235	5	A87606 Sequence 3
14	1235.4	98.6	1228	5	A87608 Sequence 5
15	1235.4	98.6	1228	5	A87608 Sequence 5
16	1235.4	97.6	1245	5	A87605 Sequence 2
17	1235.4	97.6	1245	5	A87605 Sequence 2
18	1235.4	97.6	1245	5	A87605 Sequence 2
19	1235.4	97.6	1245	5	A87605 Sequence 2
20	1235.4	97.6	1245	5	A87605 Sequence 2
21	1235.4	97.6	1245	5	A87605 Sequence 2
22	1235.4	97.6	1245	5	A87605 Sequence 2
23	1235.4	97.6	1245	5	A87605 Sequence 2
24	1235.4	97.6	1245	5	A87605 Sequence 2
25	1235.4	97.6	1245	5	A87605 Sequence 2
26	1235.4	97.6	1245	5	A87605 Sequence 2
27	1235.4	97.6	1245	5	A87605 Sequence 2
28	1235.4	97.6	1245	5	A87605 Sequence 2
29	1235.4	97.6	1245	5	A87605 Sequence 2
30	1235.4	97.6	1245	5	A87605 Sequence 2
31	1235.4	97.6	1245	5	A87605 Sequence 2
32	1235.4	97.6	1245	5	A87605 Sequence 2
33	1235.4	97.6	1245	5	A87605 Sequence 2
34	1235.4	97.6	1245	5	A87605 Sequence 2
35	1235.4	97.6	1245	5	A87605 Sequence 2
36	1235.4	97.6	1245	5	A87605 Sequence 2
37	1235.4	97.6	1245	5	A87605 Sequence 2
38	1235.4	97.6	1245	5	A87605 Sequence 2
39	1235.4	97.6	1245	5	A87605 Sequence 2
40	1235.4	97.6	1245	5	A87605 Sequence 2
41	1235.4	97.6	1245	5	A87605 Sequence 2
42	1235.4	97.6	1245	5	A87605 Sequence 2
43	1235.4	97.6	1245	5	A87605 Sequence 2
44	1235.4	97.6	1245	5	A87605 Sequence 2
45	1235.4	97.6	1245	5	A87605 Sequence 2

ALIGNMENTS

```

RESULT 1
LOCUS A87610 1229 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent W09836089.
ACCESSION A87610
VERSION A87610.1 GI:6736250
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1229)
AUTHORS Flohe, L. and Singh, M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A-20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
Source 1..1229
Location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t 1 others
ORIGIN
Query Match 99.9%; Score 1228; DB 5; Length 1229;
Best Local Similarity 100.0%; Pred. No. 2.7e-167;
Matches 1229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTTGAGATTATCGAATCTTCTTCACTGAGCGTACAGTATCGAGGGGTATC 60
DB 1 ATCTTGAGATTATCGAATCTTCTTCACTGAGCGTACAGTATCGAGGGGTATC 60
QY 61 ATGCGCGTGGGATTCGACCGAGACCAAAACAAGATTCCGGGTGGCATACCCCG 120
DB 61 ATGCGCGTGGGATTCGACCGAGACCAAAACAAGATTCCGGGTGGCATACCCCG 120
QY 121 GCGGCGTGGGAGTAACCGCGTGTGGCATGAGTGTCTATCCAGGCGAGTCCGGA 180
DB 121 GCGGCGTGGGAGTAACCGCGTGTGGCATGAGTGTCTATCCAGGCGAGTCCGGA 180
QY 181 GAGGCGTGGGATTCAGCGAGCGGATTTCAAGGGCGGCACTGTCGGCGAC 240
DB 181 GAGGCGTGGGATTCAGCGAGCGGATTTCAAGGGCGGCACTGTCGGCGAC 240
QY 241 GCCGACAGGTGTGGCGGAGCGTATTATTGCTCAAGGTCAAGAACCGATAGCGCG 300
DB 241 GCCGACAGGTGTGGCGGAGCGTATTATTGCTCAAGGTCAAGAACCGATAGCGCG 300
QY 301 GAATACGCGCGCTCGACAGCGGGGATCTTGTACGCTTCTTGATTTGGCCGCTAC 360
DB 301 GAATACGCGCGCTCGACAGCGGGGATCTTGTACGCTTCTTGATTTGGCCGCTAC 360
QY 361 GTGCTTGACCGATCGATTTGATTCGGCACCACTCAATTCCTACAGACCGTCC 420
DB 361 GTGCTTGACCGATCGATTTGATTCGGCACCACTCAATTCCTACAGACCGTCC 420
QY 421 AGACCGCGAAGGCGCACTACCCCTGCTTCCCGCATAGGGAATGCCGCTGACTCG 480
DB 421 AGACCGCGAAGGCGCACTACCCCTGCTTCCCGCATAGGGAATGCCGCTGACTCG 480
QY 481 CCGCCAGCTTGGCGCTTACCACTGATGCGAACCCAGGGGCGCGGCTGTGCTGATGG 540
DB 481 CCGCCAGCTTGGCGCTTACCACTGATGCGAACCCAGGGGCGCGGCTGTGCTGATGG 540
QY 541 GCGGGGTCCCGGCGTTCGAACCGCGGAGCTGCTGATGCGGCGCGGACCGCGGCT 600
DB 541 GCGGGGTCCCGGCGTTCGAACCGCGGAGCTGCTGATGCGGCGCGGACCGCGGCT 600
QY 601 ACAACGAGCCCGCATCGCAACGGGATGGCGCGGCGGATACGGTTTACATCAACA 660
DB 601 ACAACGAGCCCGCATCGCAACGGGATGGCGCGGCGGATACGGTTTACATCAACA 660
QY 661 TCGACAACCTTGGCACTCGACGCGGAGTTCTGGGCGCGGATTCACACTCGCTACTCAT 720

```

```

DB 661 TCGACAACCTTGGCACTCGACGCGGAGTTCTGGGCGCGGATTCACACTCGCTACTCAT 720
QY 721 CCGGCTTCAGAGCTCGAGGGTCCGCTCAACGTCGCCGACCTGTGATTTGGGCGCTCTGG 780
DB 721 CCGGCTTCAGAGCTCGAGGGTCCGCTCAACGTCGCCGACCTGTGATTTGGGCGCTCTGG 780
QY 781 TGCCAGCGCGCAAGGCGCAATTAATCTGGAATTCATCTTGGCGCATATGAACCGAG 840
DB 781 TGCCAGCGCGCAAGGCGCAATTAATCTGGAATTCATCTTGGCGCATATGAACCGAG 840
QY 841 GTGGGCTAGTGTGATATATGACCATCGACCGAGGCGGCTGTTTGAAGGCTCAGACCGA 900
DB 841 GTGGGCTAGTGTGATATATGACCATCGACCGAGGCGGCTGTTTGAAGGCTCAGACCGA 900
QY 901 CCACCTACGACACCGGACGCTGCGGCTGCACGACACGCTGTTTACTGCTGGCCAAACA 960
DB 901 CCACCTACGACACCGGACGCTGCGGCTGCACGACACGCTGTTTACTGCTGGCCAAACA 960
QY 961 TGCCGCGCTGGTGGCGCAAGCGTACGCTACGGGCTGACCAACGCGAGATGCGGTATG 1020
DB 961 TGCCGCGCTGGTGGCGCAAGCGTACGCTACGGGCTGACCAACGCGAGATGCGGTATG 1020
QY 1021 TGCTGAGCTTCCCGACCATGCTGGCGGGCGGCTGCGCGTGAATCCGCACTAGCCA 1080
DB 1021 TGCTGAGCTTCCCGACCATGCTGGCGGGCGGCTGCGCGTGAATCCGCACTAGCCA 1080
QY 1081 AAGGCTTTGAGCGACGACGAGGCGGCTTACTGTCGCAAGCGGTGGCCACCGACTGGGG 1140
DB 1081 AAGGCTTTGAGCGACGACGAGGCGGCTTACTGTCGCAAGCGGTGGCCACCGACTGGGG 1140
QY 1141 TGCCGCTACCGAGCGCGGCGGCGTGGCTGACCTCTGCGCGGCTCGTTACGCGGAG 1200
DB 1141 TGCCGCTACCGAGCGCGGCGGCGTGGCTGACCTCTGCGCGGCTCGTTACGCGGAG 1200
QY 1201 ACACGTCNGAGTAAGGGAAGCATGATG 1229
DB 1201 ACACGTCNGAGTAAGGGAAGCATGATG 1229

RESULT 2
LOCUS A89750 1229 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent W09832862.
ACCESSION A89750
VERSION A89750.1 GI:6738284
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1229)
AUTHORS Flohe, L. and Singh, M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A-30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
Source 1..1229
Location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t 1 others
ORIGIN
Query Match 99.9%; Score 1228; DB 5; Length 1229;
Best Local Similarity 100.0%; Pred. No. 2.7e-167;
Matches 1229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTTGAGATTATCGAATCTTCTTCACTGAGCGTACAGTATCGAGGGGTATC 60
DB 1 ATCTTGAGATTATCGAATCTTCTTCACTGAGCGTACAGTATCGAGGGGTATC 60
QY 61 ATGCGCGTGGGATTCGACCGAGACCAAAACAAGATTCCGGGTGGCATACCCCG 120

```

```

Db 61 ATGGCGGTGGTATTCGACCGAGAACCAAAACAAAGATTCGGGGTGGCATACCCCG 120
Qy 121 GCCGCGTGGCGAAMCTAACCCGTGTGGCATAGAGTCTCATCCAGCAGTCCGGA 180
Db 121 GCCGCGTGGCGAAMCTAACCCGTGTGGCATAGAGTCTCATCCAGCAGTCCGGA 180
Qy 181 GAGGGGTGGCGATACCGACGCGGATTCAGAGCGGCGGCGCACTGGTGGCGAC 240
Db 181 GAGGGGTGGCGATACCGACGCGGATTCAGAGCGGCGGCGCACTGGTGGCGAC 240
Qy 241 GCCGACAGAGTGGGCGCAGCCTGATTTATGCTCAAGATCAAAAGACGATAGCGCG 300
Db 241 GCCGACAGAGTGGGCGCAGCCTGATTTATGCTCAAGATCAAAAGACGATAGCGCG 300
Qy 301 GAATACGCGCGCTGGCAGACGCGGCGATTTGTTCAGCTTTTCATTTGGCCGCTCAC 360
Db 301 GAATACGCGCGCTGGCAGACGCGGCGATTTGTTCAGCTTTTCATTTGGCCGCTCAC 360
Qy 361 GTGCTTGACAGCATGGCTGTGGATTCGGGACCAACGTCATGCTAGAGACGCTCC 420
Db 361 GTGCTTGACAGCATGGCTGTGGATTCGGGACCAACGTCATGCTAGAGACGCTCC 420
Qy 421 AACACGCGCAAGGCGCAGTACCTGCTTGGCCGATGAGCGAAGTGGCGGTGACTCG 480
Db 421 AACACGCGCAAGGCGCAGTACCTGCTTGGCCGATGAGCGAAGTGGCGGTGACTCG 480
Qy 481 CGGCGCGAGTGGCGCTTACCACTGATGCGAAACCAAGGGGCGCGGTGTGATGG 540
Db 481 CGGCGCGAGTGGCGCTTACCACTGATGCGAAACCAAGGGGCGCGGTGTGATGG 540
Qy 541 GGGGGGTGGCGCGCTGCAACCGCGGAGTGTGTGATCGCGCGCGCGCGCGCGCT 600
Db 541 GGGGGGTGGCGCGCTGCAACCGCGGAGTGTGTGATCGCGCGCGCGCGCGCGCT 600
Qy 601 ACAACGACCGCGCATCGCAACGAGGCGCGACCTTACGAGTTCAGATCAACA 660
Db 601 ACAACGACCGCGCATCGCAACGAGGCGCGACCTTACGAGTTCAGATCAACA 660
Qy 661 TCGACAACTTGGGCACTGACGCGCGGAGTTCGCGCGGATGCACTCGCTACTCAT 720
Db 661 TCGACAACTTGGGCACTGACGCGCGGAGTTCGCGCGGATGCACTCGCTACTCAT 720
Qy 721 CGGCGTACAGAGTCCGAGGTCGCTTAAACGTCGCACTGTGGATGGGGCGCTCTG 780
Db 721 CGGCGTACAGAGTCCGAGGTCGCTTAAACGTCGCACTGTGGATGGGGCGCTCTG 780
Qy 781 TCGCAGGCGCAAGGACCAATATGCTCGAATTCATTGTGCGGCAATGAAACAG 840
Db 781 TCGCAGGCGCAAGGACCAATATGCTCGAATTCATTGTGCGGCAATGAAACAG 840
Qy 841 GTGCGGTACTGTGATATAGCATGACAGGCGGCTGTTTGAAGGCTCAGACCGA 900
Db 841 GTGCGGTACTGTGATATAGCATGACAGGCGGCTGTTTGAAGGCTCAGACCGA 900
Qy 901 CCACCTACAGACCGGAGGATTCGCGTGCAGACAGCTGTTTACTGCTGGGCAACA 960
Db 901 CCACCTACAGACCGGAGGATTCGCGTGCAGACAGCTGTTTACTGCTGGGCAACA 960
Qy 961 TCGCGCGCTGCTGCGGAGAGCTGACCTACGCGTGAACAGGAGAGATCCGTATG 1020
Db 961 TCGCGCGCTGCTGCGGAGAGCTGACCTACGCGTGAACAGGAGAGATCCGTATG 1020
Qy 1021 TCGCTGAGGTTGGCGACATGCTGTGGGCGGCGGTGCGGTGCAATCCGGCCTAGCA 1080
Db 1021 TCGCTGAGGTTGGCGACATGCTGTGGGCGGCGGTGCGGTGCAATCCGGCCTAGCA 1080
Qy 1081 AAGGTGTTTCAGAGCGAGGAGGCGTACTGTCGAAAGGGGCGCACGACCTGGGG 1140
Db 1081 AAGGTGTTTCAGAGCGAGGAGGCGTACTGTCGAAAGGGGCGCACGACCTGGGG 1140
Qy 1141 TCGCGTTTCAACGAGCGCGCAGGCTGTGGCTGACTCGCGCTGCTTACGCGGAG 1200
Db 1141 TCGCGTTTCAACGAGCGCGCAGGCTGTGGCTGACTCGCGCTGCTTACGCGGAG 1200

```

```

Qy 1201 ACACGTGNGAGTAAGGGAAGCGATGATG 1229
Db 1201 ACACGTGNGAGTAAGGGAAGCGATGATG 1229

RESULT 3
A87611 1235 bp DNA PAT 22-JAN-2000
LOCUS A87611
DEFINITION Sequence 8 from Patent WO9836089.
ACCESSION A87611
VERSION A87611.1 GI:6736251
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1235)
AUTHORS Flohe L. and Singh M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
source
location/Qualifiers
1..1235
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 394 c 385 g 220 t
ORIGIN

Query Match 99.9% Score 1228; DB 5; Length 1235;
Best Local Similarity 99.9%; Pred. No. 2.7e-167;
Matches 1228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATCTTGC GATTATATCGAATTTCTTACACATGAGCGTACAGTATCGAGGGGTAATC 60
Db 1 ATCTTGC GATTATATCGAATTTCTTACACATGAGCGTACAGTATCGAGGGGTAATC 60
Qy 61 ATGCGGTGGTATTCGACCGGAGACCAAAACAAAGAAATTCGGGTGCCATCACCCCG 120
Db 61 ATGCGGTGGTATTCGACCGGAGACCAAAACAAAGAAATTCGGGTGCCATCACCCCG 120
Qy 121 GCGGGGTGGCGGAACTAACCCGCTGGGCGCATGAGGTGCTATCCAGGAGTCCCGGA 180
Db 121 GCGGGGTGGCGGAACTAACCCGCTGGGCGCATGAGGTGCTATCCAGGAGTCCCGGA 180
Qy 181 GAGGGTCCGCTATACCGCAGCGGATTTCAAGGGGCGAGCGCGCAATGGTCCGACAC 240
Db 181 GAGGGTCCGCTATACCGCAGCGGATTTCAAGGGGCGAGCGCGCAATGGTCCGACAC 240
Qy 241 GCGGACAGAGTGGGCGCAGCGCTGATTTATGCTCAAGGTCAAGAACCGATAGCGGG 300
Db 241 GCGGACAGAGTGGGCGCAGCGCTGATTTATGCTCAAGGTCAAGAACCGATAGCGGG 300
Qy 301 GAATACGCGCGCTGGAGACGCGGATGTTGACGTTCTGCAATTTGGCGGGTCAAC 360
Db 301 GAATACGCGCGCTGGAGACGCGGATGTTGACGTTCTGCAATTTGGCGGGTCAAC 360
Qy 361 GTGCTTGACAGATGGTGTGGATTCGGGACCAACGTCGCAATTTGGCTACGAGACCTTC 420
Db 361 GTGCTTGACAGATGGTGTGGATTCGGGACCAACGTCGCAATTTGGCTACGAGACCTTC 420
Qy 421 AGACGCGCGAAGGCGCAGTACCTGCTTGGCCCGATGAGCAAGTGGCGGTGCACTCG 480
Db 421 AGACGCGCGAAGGCGCAGTACCTGCTTGGCCCGATGAGCAAGTGGCGGTGCACTCG 480
Qy 481 CGGCGCGAGTGGCGCTTACCACTGATGCGAACCAGAGGGGCGCGGTGTGCTGATGG 540
Db 481 CGGCGCGAGTGGCGCTTACCACTGATGCGAACCAGAGGGGCGCGGTGTGCTGATGG 540
Qy 541 GCGGGGTGGCGCGCTGCAACCGGCGAGCTGTGGTGTATCGGCGCGCGACCGCGGCT 600
Db 541 GCGGGGTGGCGCGCTGCAACCGGCGAGCTGTGGTGTATCGGCGCGCGACCGCGGCT 600

```

```

Oy 601 ACAACGACCCCGCATCGCAACGCGATGAGGCGGACCGCTTACGTTCTAGACATCAACA 660
    |||||||
Db 601 ACAACGACCCCGCATCGCAACGCGATGAGGCGGACCGCTTACGTTCTAGACATCAACA 660
Oy 661 TCGACAACTTGGGCACTCGACGCGGATTCGCGGCGGATCGACACTGCTACTCAT 720
    |||||||
Db 661 TCGACAACTTGGGCACTCGACGCGGATTCGCGGCGGATCGACACTGCTACTCAT 720
Oy 721 CGGCTTACGAGCTCGAGGCTGCGCTCAAAACGTCGCGACTGTTGATTTGGGCGGCTCTGG 780
    |||||||
Db 721 CGGCTTACGAGCTCGAGGCTGCGCTCAAAACGTCGCGACTGTTGATTTGGGCGGCTCTGG 780
Oy 781 TCGCAGGCGCCAAAGCACCACCAATTAGTCTCGAATTCATCTTCGCGCATATGAACACAG 840
    |||||||
Db 781 TCGCAGGCGCCAAAGCACCACCAATTAGTCTCGAATTCATCTTCGCGCATATGAACACAG 840
Oy 841 GTGGGCTACTGCTGATATAGCCATCGACAGGCGGCTGTTTCGAAAGCTCAGACCGA 900
    |||||||
Db 841 GTGGGCTACTGCTGATATAGCCATCGACAGGCGGCTGTTTCGAAAGCTCAGACCGA 900
Oy 901 CCACCTACGACACCCGACGCTGCGGCTGACAGACGCTGTTTACTGTCGAGGCGGATC 960
    |||||||
Db 901 CCACCTACGACACCCGACGCTGCGGCTGACAGACGCTGTTTACTGTCGAGGCGGATC 960
Oy 961 TCGCCGCTCGGTGCGCGAAGAGAGCTGACCTACGCGCTGACCAACGCGAGATCCGCTATG 1020
    |||||||
Db 961 TCGCCGCTCGGTGCGCGAAGAGAGCTGACCTACGCGCTGACCAACGCGAGATCCGCTATG 1020
Oy 1021 TCGCTGAGCTTCCGACCATGCTGCGGCGGCGCTGCGGCTGCAATCCGAGCTAGCCA 1080
    |||||||
Db 1021 TCGCTGAGCTTCCGACCATGCTGCGGCGGCGCTGCGGCTGCAATCCGAGCTAGCCA 1080
Oy 1081 AAGCTCTTCGACGACGAGGAGGCGGCTTACTGTCGAAACGCGGACCGACGCTGCGG 1140
    |||||||
Db 1081 AAGCTCTTCGACGACGAGGAGGCGGCTTACTGTCGAAACGCGGACCGACGCTGCGG 1140
Oy 1141 TCGCCTTACCGACGACGCGGACGCTGCTGAGCTCTGCGCGCTGTTACGCGGAC 1200
    |||||||
Db 1141 TCGCCTTACCGACGACGCGGACGCTGCTGAGCTCTGCGCGCTGTTACGCGGAC 1200
Oy 1201 ACACGTCGAGTAAAGGAGGAGGATGATG 1229
    |||||||
Db 1201 ACACGTCGAGTAAAGGAGGAGGATGATG 1229

RESULT 4
LOCUS A89751 1235 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 8 from Patent WO9832862.
ACCESSION A89751
VERSION A89751.1 GI:6738285
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1235)
AUTHORS Flohe, L. and Singh, M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNML Patent: WO 9832862-A 30-JUL-1998:
FEATRES FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
SOURCE 1. .1235
        /organism="unidentified"
        /db_xref="taxon:32644"
BASE COUNT 236 a 394 c 385 g 220 t
ORIGIN
Query Match 99.9%; Score 1228; DB 5; Length 1235;
Best Local Similarity 99.9%; Pred. No. 2,7e-167;
Matches 1228; Conservative 0; Mismatches 1; Indels 0; Gaps 0:
Oy 1 ATCTTGACGATTATTCGAACTTTCTTCACTGAAAGCGTACGATGAGAGGGGTAATC 60

```

```

Db 1 ATCTTGACGATTATTCGAACTTTCTTCACTGAAAGCTACGATGAGAGGGGTATC 60
    |||||||
Oy 61 ATGCGCGTGGTATTCGACGAGACCAAAACAAAGAAATTCGCGG;GGCCATCACCCG 120
    |||||||
Db 61 ATGCGCGTGGTATTCGACGAGACCAAAACAAAGAAATTCGCGG;GGCCATCACCCG 120
Oy 121 GCGGCGTGGGGAATTAACCCGTCGTGGCATAGGTGCTACTTCAGGCAAGAGTGGCGGA 180
    |||||||
Db 121 GCGGCGTGGGGAATTAACCCGTCGTGGCATAGGTGCTACTTCAGGCAAGAGTGGCGGA 180
Oy 181 GAGGCGTGGCTATCACCGACGCGGATTTCAAGGCGGACGCGCAACTGGTGGCAC 240
    |||||||
Db 181 GAGGCGTGGCTATCACCGACGCGGATTTCAAGGCGGACGCGCAACTGGTGGCAC 240
Oy 241 GCGGACAGGTGTGGGCGGACGCTGATTTATGCTCAAGGTCAAGAACGAGATAGCGCG 300
    |||||||
Db 241 GCGGACAGGTGTGGGCGGACGCTGATTTATGCTCAAGGTCAAGAACGAGATAGCGCG 300
Oy 301 GAATACGCGCGCTGCGACAGGCGGATCTGTACAGTTCTTCATTTGGCGCGCTAC 360
    |||||||
Db 301 GAATACGCGCGCTGCGACAGGCGGATCTGTACAGTTCTTCATTTGGCGCGCTAC 360
Oy 361 GTGCTTGACCCGATGCTGTTGGATTCCGGACACGCTCAATTCCTTACGAGACGCTCC 420
    |||||||
Db 361 GTGCTTGACCCGATGCTGTTGGATTCCGGACACGCTCAATTCCTTACGAGACGCTCC 420
Oy 421 AGACCGCGGAAGGCGGCACTACCCCTGCTGCGCGGATGAGGAGAGTGGCGGTGACTCG 480
    |||||||
Db 421 AGACCGCGGAAGGCGGCACTACCCCTGCTGCGCGGATGAGGAGAGTGGCGGTGACTCG 480
Oy 481 CCGCCCAAGTTGGGCGCTTACCACTGATGCGAACCAAGGGGCGCGGCGTGGTATGG 540
    |||||||
Db 481 CCGCCCAAGTTGGGCGCTTACCACTGATGCGAACCAAGGGGCGCGGCGTGGTATGG 540
Oy 541 GCGGCGTCCCGGCGGTGGAACCGGCGCGCGTGGTATGGCGCGCGCGCT 600
    |||||||
Db 541 GCGGCGTCCCGGCGGTGGAACCGGCGCGCGTGGTATGGCGCGCGCGCT 600
Oy 601 ACAACGACCCCGCATCGCAACGCGATGAGGCGGACCGCTTACGCTTACATCAACA 660
    |||||||
Db 601 ACAACGACCCCGCATCGCAACGCGATGAGGCGGACCGCTTACGCTTACATCAACA 660
Oy 661 TCGACAACTTCCGCACTCGACGCGGATTCGCGGCGGATTCGACACTGCTACTCAT 720
    |||||||
Db 661 TCGACAACTTCCGCACTCGACGCGGATTCGCGGCGGATTCGACACTGCTACTCAT 720
Oy 721 CGGCTTACGAGCTCGAGGCTGCGCTCAAAACGTCGCGACTGTTGATTTGGGCGGCTCTGG 780
    |||||||
Db 721 CGGCTTACGAGCTCGAGGCTGCGCTCAAAACGTCGCGACTGTTGATTTGGGCGGCTCTGG 780
Oy 781 TCGCAGGCGCCAAAGCACCACCAATTAGTCTCGAATTCATCTTCGCGCATATGAACACAG 840
    |||||||
Db 781 TCGCAGGCGCCAAAGCACCACCAATTAGTCTCGAATTCATCTTCGCGCATATGAACACAG 840
Oy 841 GTGGGCTACTGCTGATATAGCCATCGACAGGCGGCTGTTTCGAAAGCTCAGACCGA 900
    |||||||
Db 841 GTGGGCTACTGCTGATATAGCCATCGACAGGCGGCTGTTTCGAAAGCTCAGACCGA 900
Oy 901 CCACCTACGACACCCGACGCTGCGGCTGACAGACGCTGTTTACTGTCGAGGCGGATC 960
    |||||||
Db 901 CCACCTACGACACCCGACGCTGCGGCTGACAGACGCTGTTTACTGTCGAGGCGGATC 960
Oy 961 TCGCCGCTCGGTGCGCGAAGAGAGCTGACCTACGCGCTGACCAACGCGAGATCCGCTATG 1020
    |||||||
Db 961 TCGCCGCTCGGTGCGCGAAGAGAGCTGACCTACGCGCTGACCAACGCGAGATCCGCTATG 1020
Oy 1021 TCGCTGAGCTTCCGACCATGCTGCGGCGGCGCTGCGGCTGCAATCCGAGCTAGCCA 1080
    |||||||
Db 1021 TCGCTGAGCTTCCGACCATGCTGCGGCGGCGCTGCGGCTGCAATCCGAGCTAGCCA 1080
Oy 1081 AAGCTCTTCGACGACGAGGAGGCGGCTTACTGTCGAAACGCGGACCGACGCTGCGG 1140
    |||||||

```

Db 1081 AAGTCTTTCGACGACGAGGAGGCGCTACTGTCCGAAAGGGGTGGCCAGCAGCTGGGG 1140  
QY 1141 TCCGCTTACAGAGCCGCGACAGGTGCTGCTGACTCGGCGCGCTGCTTACGCGGAGC 1200  
Db 1141 TCCGCTTACAGAGCCGCGACAGGTGCTGCTGACTCGGCGCGCTGCTTACGCGGAGC 1200  
QY 1201 ACACGTGAGTAAAGGAGGATGATG 1229  
Db 1201 ACACGTGAGTAAAGGAGGATGATG 1229  
RESULT 5  
LOCUS A87609 1235 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 6 from Patent WO9836089.  
ACCESSION A87609  
VERSION A87609.1 GI:6736249  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1235)  
AUTHORS Flohe, L. and Singh, M.  
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
FEATURES  
source Location/Qualifiers  
1..1235  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 235 a 395 c 384 g 220 t 1 others  
ORIGIN

Query Match 99.7% Score 1225.4; DB 5; Length 1235;  
Best Local Similarity 99.8%; Pred. No. 6.4e-167;  
Matches 1226; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ACTTGCAGATTAAATGAGACTTCTTACACGTAAGCGTACAGTATCGAGAGGGTATC 60  
Db 1 ACTTGCAGATTAAATGAGACTTCTTACACGTAAGCGTACAGTATCGAGAGGGTATC 60  
QY 61 ATGCGGCTGCTATTCGACGACGACAAACAAACAAAGTTCGGGTGGCCATCAACCCG 120  
Db 61 ATGCGGCTGCTATTCGACGACGACAAACAAACAAAGTTCGGGTGGCCATCAACCCG 120  
QY 121 GCGGCGTGGCGGAACTAACCCGCTGCTGCGCATGAGTGTCTATCCAGCAGGTGCGGA 180  
Db 121 GCGGCGTGGCGGAACTAACCCGCTGCTGCGCATGAGTGTCTATCCAGCAGGTGCGGA 180  
QY 181 GAGGCTGGCTATACGACGCGGATTTCAAGGGGCGGCGGCACTGGTGGCGACC 240  
Db 181 GAGGCTGGCTATACGACGCGGATTTCAAGGGGCGGCGGCACTGGTGGCGACC 240  
QY 241 GCCGACAGGTGCGGCGGACGCTGATTTATGCTCAAGTCAAAAGAACCGATAGCGGG 300  
Db 241 GCCGACAGGTGCGGCGGACGCTGATTTATGCTCAAGTCAAAAGAACCGATAGCGGG 300  
QY 301 GAATACGGGCGGCTGGGACAGGGGCGATCTGTTCACGTTTCTGCAATTTGGCGGCTAC 360  
Db 301 GAATACGGGCGGCTGGGACAGGGGCGATCTGTTCACGTTTCTGCAATTTGGCGGCTAC 360  
QY 361 GAGCTGACAGGATGCTGTTGATTCGGGACGACGTCATTTGCTTGCAGAGCGTGC 420  
Db 361 GAGCTGACAGGATGCTGTTGATTCGGGACGACGTCATTTGCTTGCAGAGCGTGC 420  
QY 421 AGACGCGCAGAGCGGACGACGCTGCTTGGCCGATGAGCGAAGTGGCGGCTGACTGC 480  
Db 421 AGACGCGCAGAGCGGACGACGCTGCTTGGCCGATGAGCGAAGTGGCGGCTGACTGC 480  
QY 481 CCGCCAGAGTGGCGCTTACACATGATGGAACCCAAAGGGGCGCGGTGTCTATGG 540  
Db 481 CCGCCAGAGTGGCGCTTACACATGATGGAACCCAAAGGGGCGCGGTGTCTATGG 540

QY 541 GCGGGGTGCGGCGCTGACACGCGCGGACGTCGTGTGATGCGGCGCGGACCGCGGCT 600  
Db 541 GCGGGGTGCGGCGCTGACACGCGCGGACGTCGTGTGATGCGGCGCGGACCGCGGCT 600  
QY 601 ACACGACGCGCGGATTCGCAACGAGCATGAGGCGGACGTTACGTTTACATCAACA 660  
Db 601 ACACGACGCGCGGATTCGCAACGAGCATGAGGCGGACGTTACGTTTACATCAACA 660  
QY 661 TCGACAAACTTGGGCACTCGACGCGGAGTTCTGCGCGGATCCACACAGCTACTCAT 720  
Db 661 TCGACAAACTTGGGCACTCGACGCGGAGTTCTGCGCGGATCCACACAGCTACTCAT 720  
QY 721 CGGCTACGAGCTCGAGGTCGCGTCACAAACGTCGCGGACCTGCTGATTTGGGCGGCTGG 780  
Db 721 CGGCTACGAGCTCGAGGTCGCGTCACAAACGTCGCGGACCTGCTGATTTGGGCGGCTGG 780  
QY 781 TCGCAGGCGCCAAAGGACACCAATTAATGCTGCAATTCATTTGCGGATATGAACAG 840  
Db 781 TCGCAGGCGCCAAAGGACACCAATTAATGCTGCAATTCATTTGCGGATATGAACAG 840  
QY 841 GTGCGCTACTGTGATATATGATCATGACACGAGCGGCTGTTTGAAGGCTCAGACCGA 900  
Db 841 GTGCGCTACTGTGATATATGATCATGACACGAGCGGCTGTTTGAAGGCTCAGACCGA 900  
QY 901 CCACCTACGACACCGGACGCTGCGCGTGCACAGACAGCTGTTTACTGCGGCGGGAACA 960  
Db 901 CCACCTACGACACCGGACGCTGCGCGTGCACAGACAGCTGTTTACTGCGGCGGGAACA 960  
QY 961 TCGCGCGCTGCTGCGGAGAGAGTGCACCTACGCGCTGACCAACGAGATGCGGTATG 1020  
Db 961 TCGCGCGCTGCTGCGGAGAGAGTGCACCTACGCGCTGACCAACGAGATGCGGTATG 1020  
QY 1021 TCGTGCAGCTTCCGACCATGCTGCTGCGGCGGCTGCGGCTGCAATCGGCACTAGCCA 1080  
Db 1021 TCGTGCAGCTTCCGACCATGCTGCTGCGGCGGCTGCGGCTGCAATCGGCACTAGCCA 1080  
QY 1081 AAGGCTTTTCGACGACGAGTGGGCTTACTGTCCGAACGGGTGTCACCGACCTGGGG 1140  
Db 1081 AAGGCTTTTCGACGACGAGTGGGCTTACTGTCCGAACGGGTGTCACCGACCTGGGG 1140  
QY 1141 TCGCGTTACCGAGCGCGGACGCTGCTGCGCTGACTGCTGCGCGCTGTTACGCGGAGC 1200  
Db 1141 TCGCGTTACCGAGCGCGGACGCTGCTGCGCTGACTGCTGCGCGCTGTTACGCGGAGC 1200  
QY 1201 ACACGTGAGTAAAGGAGGATGATG 1229  
Db 1201 ACACGTGAGTAAAGGAGGATGATG 1229  
RESULT 6  
LOCUS A89749 1235 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 6 from Patent WO9832862.  
ACCESSION A89749  
VERSION A89749.1 GI:6738283  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1235)  
AUTHORS Flohe, L. and Singh, M.  
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
FEATURES  
source Location/Qualifiers  
1..1235  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 235 a 395 c 384 g 220 t 1 others  
ORIGIN

Query Match 99.7%; Score 1225.4; DB 5; Length 1235;  
 Best Local Similarity 99.8%; Pred. No. 6.4e-167;  
 Matches 1226; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 ATCTTCAGATTAAATGAACTTCTTACACATGAGCTAGACGTATCGAGAGGGGTATC 60
  |||||||
DB 1 ATCTTCAGATTAAATGAACTTCTTACACATGAGCTAGACGTATCGAGAGGGGTATC 60
QY 61 ATGCGGTGCGGTATTCGACCGAGACCAAAACAAAGAAATTCGGGGTGGCATCACCCCG 120
  |||||||
DB 61 ATGCGGTGCGGTATTCGACCGAGACCAAAACAAAGAAATTCGGGGTGGCATCACCCCG 120
QY 121 GCGGGGTGCGGGAATTAACCCGTCTGGCATGAGAGTGTCTATCCAGCAGGTGCCGA 180
  |||||||
DB 121 GCGGGGTGCGGGAATTAACCCGTCTGGCATGAGAGTGTCTATCCAGCAGGTGCCGA 180
QY 181 GAGGGCTCGGCTATACCGACGCGGATTTCAAGGCGGAGCGCCCAATGGTGCAGAC 240
  |||||||
DB 181 GAGGGCTCGGCTATACCGACGCGGATTTCAAGGCGGAGCGCCCAATGGTGCAGAC 240
QY 241 GCGGACAGGTGCGGCGGCGGACGCTGATTTATTTCTCAAGAGTCAAGAACCGATAGCGCG 300
  |||||||
DB 241 GCGGACAGGTGCGGCGGCGGACGCTGATTTATTTCTCAAGAGTCAAGAACCGATAGCGCG 300
QY 301 GAATACGCGCGCTCGGACGAGCGGCGATCTGTTACGTTCTTACGTTGGCGCGCTAC 360
  |||||||
DB 301 GAATACGCGCGCTCGGACGAGCGGCGATCTGTTACGTTCTTACGTTGGCGCGCTAC 360
QY 361 GTGCTTGCACCGATGCTGTTGGATTCGGGACACAGTCAATGGCTACAGAGACCGTCC 420
  |||||||
DB 361 GTGCTTGCACCGATGCTGTTGGATTCGGGACACAGTCAATGGCTACAGAGACCGTCC 420
QY 421 AGACCGCGGAGGAGGAGTACCCCTGCTGGCCGATGAGCGAGTCCGGGTGCACTG 480
  |||||||
DB 421 AGACCGCGGAGGAGGAGTACCCCTGCTGGCCGATGAGCGAGTCCGGGTGCACTG 480
QY 481 CGGCGGAGGTGGCGCTTACCACTGATGCGAAGCCAGGCGCGGTGTGTGATG 540
  |||||||
DB 481 CGGCGGAGGTGGCGCTTACCACTGATGCGAAGCCAGGCGCGGTGTGTGATG 540
QY 541 GGGGGGTGCGCGGCTCGAAGCGCGGAGTGTGTGATGCGCGCGCGCGCGCGGT 600
  |||||||
DB 541 GGGGGGTGCGCGGCTCGAAGCGCGGAGTGTGTGATGCGCGCGCGCGCGCGGT 600
QY 601 ACAAGCAGCGCGCATGCGCAACGGCATGGGCGGACGCTTACGCTTACAGATCAACA 660
  |||||||
DB 601 ACAAGCAGCGCGCATGCGCAACGGCATGGGCGGACGCTTACGCTTACAGATCAACA 660
QY 661 TCGACAAAATTTGGGCACTCGACGCGGAGTCTGGGGCGGATCCACTCTCTACTAT 720
  |||||||
DB 661 TCGACAAAATTTGGGCACTCGACGCGGAGTCTGGGGCGGATCCACTCTCTACTAT 720
QY 721 CGGCGTACGAGCTCGAGGAGTGGCGTCAACGTCGACCTGTGATGGGGCGGCTG 780
  |||||||
DB 721 CGGCGTACGAGCTCGAGGAGTGGCGTCAACGTCGACCTGTGATGGGGCGGCTG 780
QY 781 TCGCAGGCGCCAAAGCAGCCAAATTAATGTCGAATTCACCTTGCAGCATATGAACCG 840
  |||||||
DB 781 TCGCAGGCGCCAAAGCAGCCAAATTAATGTCGAATTCACCTTGCAGCATATGAACCG 840
QY 841 GTCGCGTACGAGTATGATGATGACATGACAGGAGGCGGCTTCAAGGCTACAGACGA 900
  |||||||
DB 841 GTCGCGTACGAGTATGATGATGACATGACAGGAGGCGGCTTCAAGGCTACAGACGA 900
QY 901 CCACCTACGACACCGGAGCTGCGCGTGCACGACAGCTGTTTACTGCGTGGGGAACA 960
  |||||||
DB 901 CCACCTACGACACCGGAGCTGCGCGTGCACGACAGCTGTTTACTGCGTGGGGAACA 960
QY 961 TCGCGCGCGCTGCGGAGAGAGTGCAGCTAGCGGCTGACCAACGAGAGTCCGCTATG 1020
  |||||||
DB 961 TCGCGCGCGCTGCGGAGAGAGTGCAGCTAGCGGCTGACCAACGAGAGTCCGCTATG 1020
QY 1021 TCGCTGAGAGTTCGCGACATGAGCTGCGGCGGCGTGCAGGTCGAATCCGGCATAGCA 1080
  |||||||

```

```

DB 1021 TCGCTGAGAGTTCGCGACATGAGCTGCGGCGGCGTGCAGGTCGAATCCGGCATAGCA 1080
  |||||||
QY 1081 AAGGCTTTTCGACGACGAGAGGGGCTTACTGTCCGAACGGGTGGCCACCGACTGGGG 1140
  |||||||
DB 1081 AAGGCTTTTCGACGACGAGAGGGGCTTACTGTCCGAACGGGTGGCCACCGACTGGGG 1140
QY 1141 TCGCGTTCGAGGAGCGCCGACGCGTGTGCTGACCTCGGCGGCTGTTACGCCGAGC 1200
  |||||||
DB 1141 TCGCGTTCGAGGAGCGCCGACGCGTGTGCTGACCTCGGCGGCTGTTACGCCGAGC 1200
QY 1201 ACAGCTCGAGAGTAAAGGAGCGATGATG 1229
  |||||||
DB 1201 ACAGCTCGAGAGTAAAGGAGCGATGATG 1229
  |||||||

RESULT 7
A87613
LOCUS A87613 1236 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 10 from Patent WO9836089.
ACCESSION A87613
VERSION A87613.1 GI:6736253
KEYWORDS
SOURCE
ORANISM unidentified.
unclassified.
REFERENCE 1 (bases 1 to 1236)
AUTHORS Flohe, L. and Singh, M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A-20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MANAVIR (DE)
FEATURES
Source location/Qualifiers
1..1236
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 395 c 385 g 220 t
ORIGIN

Query Match 98.9%; Score 1215.4; DB 5; Length 1236;
Best Local Similarity 99.8%; Pred. No. 1.7e-165;
Matches 1227; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```



```

Db 421 CAGACCCGCGACGCGACATACCCCTGCTTGCCCGCATGAGCGAAGTCCGCGGTGCACTC 480
Oy 480 GCGGCCAGGTTGGGCTTACACGATGCGAAGCCAAAGGCGCCGCGGTGCGATG 539
Db 481 GCGGCCAGGTTGGGCTTACACGATGCGAAGCCAAAGGCGCGGTGCGATG 540
Oy 540 GCGGCCGTCGCGGCTGCAACGCGCGGCGATGCTGTGATCGCGCGCGCGCGCG 599
Db 541 GCGGCCGTCGCGGCTGCAACGCGCGGCGATGCTGTGATCGCGCGCGCGCG 600
Oy 600 TACACGCGACCGCGCATCGCCAAAGCGATGCGCGCGCGCGCGCGCGCG 659
Db 601 TACACGCGACCGCGCATCGCCAAAGCGATGCGCGCGCGCGCGCGCGCG 660
Oy 660 ATGACAAATTCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 719
Db 661 ATGACAAATTCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
Oy 720 TCGGCTACGAGCTCGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 779
Db 721 TCGGCTACGAGCTCGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
Oy 780 GTGCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839
Db 781 GTGCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
Oy 840 GTGCCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 899
Db 841 GTGCCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 900
Oy 900 ACCACCTAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 959
Db 901 ACCACCTAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
Oy 960 ATGCCCCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1019
Db 961 ATGCCCCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
Oy 1020 GTGCTGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1079
Db 1021 GTGCTGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
Oy 1080 AAAGGCTTTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1139
Db 1081 AAAGGCTTTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
Oy 1140 GTGCCGTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1199
Db 1141 GTGCCGTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200
Oy 1200 CACACGTCGAGTGAAGGAGGAGGATG 1229
Db 1201 CACACGTCGAGTGAAGGAGGAGGATG 1230

```

RESULT 8  
 LOCUS A89753 1236 bp DNA  
 DEFINITION Sequence 10 from Patent WO9832862.  
 ACCESSION A89753  
 VERSION A89753.1 GI:6738287  
 KEYWORDS  
 ORGANISM  
 SOURCE  
 ORGANISM  
 1 (bases 1 to 1236)  
 FLOHE L. and Singh, M.  
 L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
 Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 Location/Qualifiers  
 1..1236  
 /organism="unidentified"

```

BASE COUNT 236 a 395 c 385 g 220 t
ORIGIN
Query Match 98.9%; Score 1215.4; DB 5; Length 1236;
Best Local Similarity 99.8%; Pred. No. 1.7e-165;
Matches 1227; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```

/db\_xref="taxon:32644"

```

Oy 1 ATCTTGACGATTATGCAACTTCTTACACTGAAAGCGTACGATGAGAGGGTAAATC 60
Db 1 ATCTTGACGATTATGCAACTTCTTACACTGAAAGCGTACGATGAGAGGGTAAATC 60
Oy 61 ATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
Db 61 ATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
Oy 121 GCGGCGTTCGCGGAACTAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
Db 121 GCGGCGTTCGCGGAACTAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
Oy 181 GAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db 181 GAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Oy 241 GCGGACGAGTGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db 241 GCGGACGAGTGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Oy 301 GAATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 359
Db 301 GAATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Oy 360 GGTGCTTGCAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 419
Db 361 GGTGCTTGCAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Oy 420 CAGACCGCGCGGAGGCGCATACCGCTGCTGCGCGCGCGCGCGCGCGCGCG 479
Db 421 CAGACCGCGCGGAGGCGCATACCGCTGCTGCGCGCGCGCGCGCGCGCGCG 480
Oy 480 GCGGCGCGAGTGTGCGGCTTACACCTGATGCGAACCGCGCGCGCGCGCG 539
Db 481 GCGGCGCGAGTGTGCGGCTTACACCTGATGCGAACCGCGCGCGCGCGCG 540
Oy 540 GCGGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 599
Db 541 GCGGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Oy 600 TACAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 659
Db 601 TACAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Oy 660 ATGACAAATTCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 719
Db 661 ATGACAAATTCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
Oy 720 TCGGCTACGAGCTCGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 779
Db 721 TCGGCTACGAGCTCGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
Oy 780 GTGCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839
Db 781 GTGCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
Oy 840 GTGCCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 899
Db 841 GTGCCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 900
Oy 900 ACCACCTAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 959
Db 901 ACCACCTAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960

```

QY 960 ATGCCCGCTGGTGGCCGAGAGCTGACCTTACGCGCTGACCAAGCAGCATGCGTAT 1019  
 DB 961 ATGCCCGCTGGTGGCCGAGAGCTGACCTTACGCGCTGACCAAGCAGCATGCGTAT 1020  
 QY 1020 GTGCTGACGCTTGGCCGACCATGGTGGGGGGGCGGCTGCGGCTGCAATTCGGCACTACCC 1079  
 DB 1021 GTGCTGACGCTTGGCCGACCATGGTGGGGGGGCGGCTGCGGCTGCAATTCGGCACTACCC 1080  
 QY 1080 AAGGCTCTTCCAGCAGCAAGGGGCGCTTACCTGCGAAGGGGTGGCCACGCACTGCGGG 1139  
 DB 1081 AAGGCTCTTCCAGCAGCAAGGGGCGCTTACCTGCGAAGGGGTGGCCACGCACTGCGGG 1140  
 QY 1140 GTGCGCTTACGAGCGCCGACGCGTGTGGCTGACCTGCGGCTGCTGTTACCGCCGAG 1199  
 DB 1141 GTGCGCTTACGAGCGCCGACGCGTGTGGCTGACCTGCGGCTGCTGTTACCGCCGAG 1200  
 QY 1200 CACAGCTGNGAGTAAAGGAGACGATGATG 1229  
 DB 1201 CACAGCTGNGAGTAAAGGAGACGATGATG 1230

## RESULT 9

LOCUS A87607 1237 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 4 from Patent WO9836089.  
 ACCESSION A87607  
 VERSION A87607.1 GI:6736247  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.

REFERENCE 1 (bases 1 to 1237)  
 FLOHE, L. and Singh, M.  
 TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 Patent: WO 9836089-A 20-AUG-1998;  
 JOURNAL FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)

FEATURES  
 Location/Qualifiers  
 1..1237

BASE COUNT 236 a 394 c 386 g 221 t  
 ORIGIN  
 1..1237  
 /organism="unidentified"  
 /db\_xref="taxon:32644"

Query Match 98.8%; Score 1213.8; DB 5; Length 1237;  
 Best Local Similarity 99.7%; Pred. No. 2.9e-165;  
 Matches 1226; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 ATCTGAGATTAAATGAACTTCTTACACTGAGCGTACAGTATCGAGGGGTATC 60  
 DB 1 ATCTGAGATTAAATGAACTTCTTACACTGAGCGTACAGTATCGAGGGGTATC 60  
 QY 61 ATGCGGTGGTATTCGACGAGACCAAAACAGCAATTCGGGTGGCATACCCCG 120  
 DB 61 ATGCGGTGGTATTCGACGAGACCAAAACAGCAATTCGGGTGGCATACCCCG 120  
 QY 121 GCGGGGTCGGGAATCAACCGTGGGCGCATGAGTGTCTATCCAGGAGTGGCGGA 180  
 DB 121 GCGGGGTCGGGAATCAACCGTGGGCGCATGAGTGTCTATCCAGGAGTGGCGGA 180  
 QY 181 GAGGGCTCGCTATCAACGAGCGGATTTCAAGGGGCGGCGCAACTGTGGCACCC 240  
 DB 181 GAGGGCTCGCTATCAACGAGCGGATTTCAAGGGGCGGCGCAACTGTGGCACCC 240  
 QY 241 GCGGACAGTGTGGGCGGCGGCGGATTTATGCTCAAGGTCAAAAGACCGATAGGGGG 300  
 DB 241 GCGGACAGTGTGGGCGGCGGCGGATTTATGCTCAAGGTCAAAAGACCGATAGGGGG 300  
 QY 301 GAATAGCGCGCTGCGACACGAGGAC-GATCTGTTCAGTCTTGTGATTTGCGCGGTCA 359  
 DB 301 GAATAGCGCGCTGCGACACGAGGACGATCTGTTCAGTCTTGTGATTTGCGCGGTCA 360  
 QY 360 CGTGTTCAGCAGCGATGCTGTGTGATTCGGGACACAGCTCAATTCGCTTACGAGCCGTC 419

DB 361 CGTCTTGCACCGATGGGTGTGATTCGGGACACCAAGCAATTCCTACAGAGACCGTC 420  
 QY 420 CAGACCGCCGAGGCGGACACTACCCCTGCTTGGCCGATGAGCAATTCGCCGGTGCATTC 479  
 DB 421 CAGACCGCCGAGGCGGACACTACCCCTGCTTGGCCGATGAGCAATTCGCCGGTGCATTC 480  
 QY 480 GCGGCCAGGTGGCGCTTACCACTGATGCGAACCAGAGGGGCGCGGTGTGTGATG 539  
 DB 481 GCGGCCAGGTGGCGCTTACCACTGATGCGAACCAGAGGGGCGCGGTGTGTGATG 540  
 QY 540 GCGGGGTGCCCGGCTGCAACCGGCGGACGCTGTGGTATGCGGCGCGGACCGCGCCG 599  
 DB 541 GCGGGGTGCCCGGCTGCAACCGGCGGACGCTGTGGTATGCGGCGCGGACCGCGCCG 600  
 QY 600 TACAAGCAGCGCGCATCGGCAACGGCATGGGCGGCGGCTTACGCTTACATCAAC 659  
 DB 601 TACAAGCAGCGCGCATCGGCAACGGCATGGGCGGCGGCTTACGCTTACATCAAC 660  
 QY 660 ATGCAAAACTTGGCAACTGCAAGCGCGGAGTTCTGGGCGGATTCACACTGCTACTCA 719  
 DB 661 ATGCAAAACTTGGCAACTGCAAGCGCGGAGTTCTGGGCGGATTCACACTGCTACTCA 720  
 QY 720 TCGGCTTACGAGCTCGAGGGGTCCGCTCAACGTGGCGGACCTGGTATGGGGCGGCTCTG 779  
 DB 721 TCGGCTTACGAGCTCGAGGGGTCCGCTCAACGTGGCGGACCTGGTATGGGGCGGCTCTG 780  
 QY 780 GTGCGAGCGCGCAAGGACCCCAATTAATGCTCGAATTCACCTTGGCGCATATGAACCA 839  
 DB 781 GTGCGAGCGCGCAAGGACCCCAATTAATGCTCGAATTCACCTTGGCGCATATGAACCA 840  
 QY 840 GTGCGGCTGCTGTGATATAGCATATGACGAGGCGGCTGCTTTCGAAGCTCAGACCG 899  
 DB 841 GTGCGGCTGCTGTGATATAGCATATGACGAGGCGGCTGCTTTCGAAGCTCAGACCG 900  
 QY 900 AACACCTTACGACCGCGGAGCTTGGCGGCTGACAGCAGCTGTTTACTGCTGAGGCAAC 959  
 DB 901 AACACCTTACGACCGCGGAGCTTGGCGGCTGACAGCAGCTGTTTACTGCTGAGGCAAC 960  
 QY 960 ATGCCCGCTGCTGCGGAGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGAT 1019  
 DB 961 ATGCCCGCTGCTGCGGAGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGAT 1020  
 QY 1020 GTGCTGAGCTTGGCGACCATGGTGGGGGCGGCTGCGGCTGGAATTCGGGCACTAGCC 1079  
 DB 1021 GTGCTGAGCTTGGCGACCATGGTGGGGGCGGCTGCGGCTGGAATTCGGGCACTAGCC 1080  
 QY 1080 AAGGCTCTTTCGACGACGAGGCGGCTTACTGTCGAAAGGCTGGCGGCTGGGG 1139  
 DB 1081 AAGGCTCTTTCGACGACGAGGCGGCTTACTGTCGAAAGGCTGGCGGCTGGGG 1140  
 QY 1140 GTGCGCTTACAGGCGCGGCGGCGGCTGCTGCTGATCTTCGGCGGCTTACGCGGAG 1199  
 DB 1141 GTGCGCTTACAGGCGCGGCGGCGGCTGCTGCTGATCTTCGGCGGCTTACGCGGAG 1200  
 QY 1200 CACAGCTGNGAGTAAAGGAGACGATGATG 1229  
 DB 1201 CACAGCTGNGAGTAAAGGAGACGATGATG 1230

## RESULT 10

LOCUS A89747 1237 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 4 from Patent WO9832862.  
 ACCESSION A89747  
 VERSION A89747.1 GI:6738281  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.

REFERENCE 1 (bases 1 to 1237)  
 FLOHE, L. and Singh, M.  
 L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM

JOURNAL Patent: WO 9832862-A 30-JUL-1998;

FLOHE LEBOLD (DE); SINGH MAHAVIR (DE)

FEATURES

source

1.1237

Location/Qualifiers

/organism="unidentified"

/db\_xref="taxon:32644"

BASE COUNT

236 a 394 c 386 g 221 t

ORIGIN

Query Match 98.8%; Score 1213.8; DB 5; Length 1237;

Best Local Similarity 99.7%; Pred. No. 2.9e-165;

Matches 1226; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 ATCTTCAGATTAAATGAACTTTCTTACACAGTGAAGCTGATGAGTATGAGAGGGTAAATC 60  
 Db 1 ATCTTCAGATTAAATGAACTTTCTTACACAGTGAAGCTGATGAGTATGAGAGGGTAAATC 60  
 QY 61 ATGGCGGTGGTATTCCGACCGAGCAAAACAAAGATTCGGGTGGCCATACCCCG 120  
 Db 61 ATGGCGGTGGTATTCCGACCGAGCAAAACAAAGATTCGGGTGGCCATACCCCG 120  
 QY 121 GCGGCGGTGCGGAACTAACCCGTCGTGGCATGAGTCTCTATCCAGCAGAGTGGCGGA 180  
 Db 121 GCGGCGGTGCGGAACTAACCCGTCGTGGCATGAGTCTCTATCCAGCAGAGTGGCGGA 180  
 QY 181 GAGGCGTGGCTATACACGACGCGGATTTCAAGGCGGCAAGCGCGCACTGTCGCGACC 240  
 Db 181 GAGGCGTGGCTATACACGACGCGGATTTCAAGGCGGCAAGCGCGCACTGTCGCGACC 240  
 QY 241 GCGGACAGGTGGGCGGAGCGGTGATTTATGCTCAAGGTCAAAAGACGAT/GCGGGG 300  
 Db 241 GCGGACAGGTGGGCGGAGCGGTGATTTATGCTCAAGGTCAAAAGACGAT/GCGGGG 300  
 QY 301 GAATACGGCGCGCTGCGACACGCGGCG-GATCTTGTTCAGCTTCTTGTGCGCGCGTCA 359  
 Db 301 GAATACGGCGCGCTGCGACACGCGGCG-GATCTTGTTCAGCTTCTTGTGCGCGCGTCA 359  
 QY 360 CGTCTTGCACAGGATGCGTGTGATTCGCGACACGCTCAATTGCTTACGACAGCCGTC 419  
 Db 360 CGTCTTGCACAGGATGCGTGTGATTCGCGACACGCTCAATTGCTTACGACAGCCGTC 419  
 QY 420 CAGACCGCGGAGGAGGCGCATACCCCTGCTTGGCCCGCATGAGGGAAGTGCAGGTCGACTC 479  
 Db 420 CAGACCGCGGAGGAGGCGCATACCCCTGCTTGGCCCGCATGAGGGAAGTGCAGGTCGACTC 479  
 QY 480 GCGCGCCAGGTTGGCGCTTACCACTGATGCGAACCACCAAGGGGCGCGGTGTGTGATG 539  
 Db 480 GCGCGCCAGGTTGGCGCTTACCACTGATGCGAACCACCAAGGGGCGCGGTGTGTGATG 539  
 QY 540 GCGGCGGTGCGCGGCGTGCACACCGCGGAGCTGTGTGATGCGGCGCGGACCGCGGCG 599  
 Db 540 GCGGCGGTGCGCGGCGTGCACACCGCGGAGCTGTGTGATGCGGCGCGGACCGCGGCG 599  
 QY 600 TACAAGCGAGCGCGCATGCGCAAGGCGGAGCGGCGGACCGTTACGGTTTACAGATCAAC 659  
 Db 600 TACAAGCGAGCGCGCATGCGCAAGGCGGAGCGGCGGACCGTTACGGTTTACAGATCAAC 659  
 QY 660 ATGACAAATCTTGGCACTGCAAGCGCGGATGTCGGCGCGGATGCTACACTGCTACTCA 719  
 Db 660 ATGACAAATCTTGGCACTGCAAGCGCGGATGTCGGCGCGGATGCTACACTGCTACTCA 719  
 QY 720 TCGGCGTACAGAGTGGAGGTGCGCTCAAAAGTGCAGACTGCTGATTTGGGGGCGGCTG 779  
 Db 720 TCGGCGTACAGAGTGGAGGTGCGCTCAAAAGTGCAGACTGCTGATTTGGGGGCGGCTG 779  
 QY 780 GTGCGAGGCGGCAAGCGACCAATTAATGCTGCAATTCCTTGTGCGCGCATATGAACA 839  
 Db 780 GTGCGAGGCGGCAAGCGACCAATTAATGCTGCAATTCCTTGTGCGCGCATATGAACA 839  
 QY 840 GGTGGGCTAGTGTGATTAAGCATGACGACGAGGGGGGCTTTGGAAGGCTACAGACCG 899  
 Db 840 GGTGGGCTAGTGTGATTAAGCATGACGACGAGGGGGGCTTTGGAAGGCTACAGACCG 899

QY 900 ACCACTTACGACCCACCGACGTTGCCGTCAGCAGACGCTGTTTACTGCTGGCGAAC 959  
 Db 901 ACCACTTACGACCCACCGACGTTGCCGTCAGCAGACGCTGTTTACTGCTGGCGAAC 960  
 QY 960 ATGCCCGCTGGTGGCGAAGCGTGAAGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGT 1019  
 Db 961 ATGCCCGCTGGTGGCGAAGCGTGAAGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGT 1020  
 QY 1020 GTGCTGAGCTTGGCGGACCATGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1079  
 Db 1021 GTGCTGAGCTTGGCGGACCATGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080  
 QY 1080 AAGGCTTTTGGCGACGACGAGGCGGCTTACTGTCCGAAGGGGTGGCCACGACCTGGGG 1139  
 Db 1081 AAGGCTTTTGGCGACGACGAGGCGGCTTACTGTCCGAAGGGGTGGCCACGACCTGGGG 1140  
 QY 1140 GTGCGTTACCGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1199  
 Db 1141 GTGCGTTACCGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
 QY 1200 CACAGTCNGAGTAAGGGAAGCGATGATG 1229  
 Db 1201 CACAGTCNGAGTAAGGGAAGCGATGATG 1230

## RESULT 11

MTV002

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REMARK

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REMARK

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REMARK

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REMARK

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REMARK

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REMARK

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,  
 Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,  
 Tekle, A., Bado, A., Brown, D., Brown, D., Chillingworth, T.,  
 Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,  
 Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,  
 Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,  
 Rajandream, M.A., Rogers, J., Rutter, J., Seeger, K., Skelton, S.,  
 Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S., and  
 Barrrell, B.G.  
 Deciphering the biology of *Mycobacterium tuberculosis* from the  
 complete genome sequence  
 Nature 393 (6685), 537-544 (1998)  
 98295987  
 Erratum: [[published erratum appears in Nature 1998 Nov  
 12:396(6707):190]]  
 2 (bases 1 to 56414)  
 Parkhill, J.  
 Direct Submission  
 Submitted (11-JUN-1998) submitted on behalf of the *Mycobacterium*  
*tuberculosis* sequencing and mapping teams, Sanger Centre, Wellcome  
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique  
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,  
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
 On Jun 27, 1998 this sequence version replaced gi:1264256.  
 Notes:  
 Details of *M. tuberculosis* sequencing at the Sanger Centre are  
 available on the World Wide Web.  
 (URL, [http://www.sanger.ac.uk/Projects/M\\_tuberculosis/](http://www.sanger.ac.uk/Projects/M_tuberculosis/)) CDS have  
 been renumbered from the original cosmid submissions but the old  
 gene designations are in brackets after the new gene numbers.  
 Gene prediction was based on a Hidden Markov Model of 18 genes  
 implemented in Tblastx (Krogh) supplemented with visual inspection  
 of positional base preference in codons, especially where there is  
 an increase in the observed/expected third position G + C.

CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

## FEATURES

```

source
    1..56414
    /organism="Mycobacterium tuberculosis"
    /strain="H37Rv"
    /db_xref="taxon:1773"
    <1..>56414
    /note="Fragment designated V002. Does not represent a physical clone"
    <1..233
    /organism="Mycobacterium tuberculosis"
    /strain="H37Rv"
    /db_xref="taxon:1773"
    /clone="Y154"
    /complement(3..527)
    /gene="recX"
    /complement(3..527)
    /gene="recY"
    /note="Rv2736c, (MTV002.01c), len: 174 aa. recX, similar to eg. RECX_PSEAE_P37860 regulatory protein recX from Pseudomonas aeruginosa (153 aa), fasta scores: opt: 161 z-score: 257.2 E(): 3.6e-07, 30.7% identity in 137 aa overlap. Overlaps and extends CDS from overlapping cosmid MTCY154.16c"
    /codon_start=1
    /transl_table=11
    /product="recX"
    /protein_id="CA15533.1"
    /db_xref="GI:2624257"
    /db_xref="SPTREMBL:O33280"
    /translation="MTVSCPSTSESEDEQAACICILTRSRRAELAQAKRGY PEDIGNRYLDLAAGLYDTDFEOMVOSRAANAASKRALAEHAKGVDDVITTVLGIDACGERAEKTLRARLRREVLDIDGDEARVSRRLVAMLRGYSOTLACEY VIAELAEERERRRV"
    /complement(493..2865)
    /gene="recA"
    /complement(493..2865)
    /gene="recA"
    /complement(493..2865)
    /note="Rv2737c, (MTV002.02c), len: 790 aa. recA, identical to REC_A_MYCYU_P26345 recA protein (790 aa). Contains self-splicing protein element (intein) from 2294 to 969 (C), similar to intein II from TR:E332317 (EMBL:Y13030) DNA-directed DNA polymerase (EC 2.7.7.7) from Thermococcus sp. (1829 aa), fasta scores: opt: 81 z-score: 235.2 E(): 6e-06, 24.6% identity in 183 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop), PS00321 recA signature, and PS00881 protein splicing signature. See Davis et al, (1992) Cell 71(2):201-210"
    /codon_start=1
    /transl_table=11
    /product="recA"
    /protein_id="CA15533.1"
    /db_xref="GI:2624258"
    /db_xref="SWISS-PROT:P26345"
    /translation="MTOPPRKALELAVAQIEKSYGKSVMRIGDEARODISYIPG SIADVAIGISGLPRGYETIEYGPSSSKTTVALAHANNAAGVAAFIDAEHALDP DYAKLGYDTSLVSOPDTEQALEIDMLIRSGALDIVIDVSAALVPRAELEGEG GDHVGLOARLMSQALRKMTGALNLTGTAIFINQLRDKIGVMFGSEPTTGKALF YASVMDVRYEETLKDGTNAVGNRTRVVKNKCLAGSTRIFDVTGTRHIEVVG RKPIHVAAKDEGLTHARVYVSMFDGRTDVIYIGIAGATVMAVTPHKHRIEYGMW AGELRKGRVAVQPRPFDFGSAPIPADHARLTGLIGDGDGVMGKTPINFINVOR ALIDYPRATATTCGAAPGCRSLAIARHPEGVADICCOAGTGYCKLAWERTIEN WFPEDTADIVGNLIRGLFESDCHWSHEQTGALRVGTTTSEQLAHQIHLIRFV GSTVADYDTQKRSYIVNGRIQSKRQVFEVRISGMONTVAFAESVPMGPRGALIO AIPETAGDGRGSOATYLAEMTDAVNLDERGVTAQEAAMTGAAGSDPRGCMKYO LGASLRDRDVALADLDDFLHDMLEELRSVIREVLTARRATFDLVEELHTL VAEGVYVNCSPFROAEFDILYKGISLESLIDMGVDOGLIRSGMWFYEEEOIG GQKEARNFLEVNADVDEIKIKENKIGICGAVVTTDPSNDGVLPAPVDF"

```

```

misc_feature
    complement(790..2112)
    /gene="recA"
    /note="recA intein"
    /complement(790..807)
    /gene="recA"
    /note="PS00881 Protein splicing signature"
    /complement(2197..2223)
    /gene="recA"
    /note="PS00321 recA signature"
    /complement(2644..2667)
    /gene="recA"
    /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
    /complement(2873..2877)
    /note="possible RBS upstream of Rv2737c"
    /complement(3247..3453)
    /gene="Rv2738c"
    /complement(3247..3453)
    /gene="Rv2738c"
    /note="Rv2738c, (MTV002.03c), len: 68 aa. Unknown, N-terminus is highly similar to the N-terminus of the upstream ORF MTV002.07c (78.4% identity in 37 aa overlap); also similar to AL020958|SC4H8.5 Streptomyces coelicolor cosmid 4H8 (64aa) opt: 185 z-score: 283.5 E(): 2.9e-08; 39.7% identity in 63 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein Rv2738c"
    /protein_id="CA15534.1"
    /db_xref="GI:2624260"
    /db_xref="SPTREMBL:O33281"
    /translation="MLAGVRLTFEHEVALHFGAAYGSVLDHYLTGFDGRSAQAI EDCVEPRDVWRAICADFDVPHDRW"
    /complement(3464..4630)
    /gene="Rv2739c"
    /complement(3464..4630)
    /gene="Rv2739c"
    /note="Rv2739c, (MTV002.04c), len: 388 aa. Probable transferase, similar to eg. TR:Q51560 (EMBL:L28170) Pseudomonas aeruginosa rhomnosyl transferase (426 aa), fasta scores: opt: 178 z-score: 226.3 E(): 1.9e-05, 25.9% identity in 425 aa overlap. Equivalent to Mycobacterium leprae protein MLCB33.02c (392 aa); fasta scores gpl294723|MLCB33.2 Mycobacterium leprae cosmid B33 opt: 2112 z-score: 2364.5 E(): 0; 80.9% identity in 388 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein Rv2739c"
    /protein_id="CA15535.1"
    /db_xref="GI:2624261"
    /db_xref="SPTREMBL:O33282"
    /translation="MRVAVVAGPDPHSHFPAIALCQFRRAADPTLTGVEKTEAR AACIDVNEIDGLATBDRLDAGARIRRAQAVLVAVPRRLAEPLTVSDYITACGG MAELUGIPWELNPPHLYLPKSGPLPGSLAAGVIRGLRDATRALVIGRSWAG LRQRAAVREIIGLPARDPGLRLIATLPALVPRDWPAPAAVVGPLFEPTDRLTA IPAGTGVVAVASTALTAGTLEVALOSTVGETVPSGSLVYSRLSGADLVYPM AVAGISQAEILTRADIVICGGHGWAKLLGVPMVYVGGGDMETANRVRGGS AVLIRPLTADALVAANEVLSPRFREARRAASVAGAADPVRVCHDALALAG"
    /complement(4674..5123)
    /gene="Rv2740"
    /complement(4674..5123)
    /gene="Rv2740"
    /note="Rv2740, (MTV002.05), len: 149 aa. Unknown, similar to M. leprae hypothetical protein TR:Q49850 (EMBL:294723) MLCB33.03 (B2235.F3.140) (178 aa), fasta scores: opt: 498 z-score: 538.9 E(): 7.3e-23, 51.6% identity in 161 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein Rv2740"
    /protein_id="CA15536.1"
    /db_xref="GI:2624262"
    /db_xref="SPTREMBL:O33283"

```

gene  
CDS

/translation="MAETLSPENPETTAIRAEVAFNLALONEDPDTVAALGDL  
VYENVGSRIKGRRTATILRMGRNVEFKIRHIGDGAVALTEPDAIILGPIRV  
QFWVGVFVDDGRITLMRDYFDYDMFKGLRLGLVALVPSLATL-  
5355..6932  
/gene="PE\_PGRS"  
5355..6932  
/gene="PE\_PGRS"  
/note="RV2741, (MTV002.06), len: 525 aa; Member of M.  
tuberculosis PE\_PGRS subfamily, similar to many eg.

Query Match 98.8%; Score 1213.8; DB 1; Length 56414;  
Best Local Similarity 99.7%; Pred. No. 1.2e-165;  
Matches 1226; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 ATCTTGACATTAATGAACTTTCTTACACTGAAAGCTATCGAGGGGTAATC 60  
DB 38201 ATCTTGACATTAATGAACTTTCTTACACTGAAAGCTATCGAGGGGTAATC 38260  
QY 61 ATGCGGTGGGTATTCGACGAGAACCAAAACAAAGATTCGGGGTGGCCATCAACCCG 120  
DB 38261 ATGCGGTGGGTATTCGACGAGAACCAAAACAAAGATTCGGGGTGGCCATCAACCCG 38320  
QY 121 GCGGCGTGGGAACTAACCCGTGTGCGCATGAGTGTCT ATCCAGCAGGTGCCGGA 180  
DB 38321 GCGGCGTGGGAACTAACCCGTGTGCGCATGAGTGTCTATCCAGCAGGTGCCGGA 38380  
QY 181 GAGGGCTGGGTATTCACGACGCGGATTTCAAAGCGGCGCACTGCTGGCCACC 240  
DB 38381 GAGGGCTGGGTATTCACGACGCGGATTTCAAAGCGGCGCACTGCTGGCCACC 38440  
QY 241 GCGGACAGGTGGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGACGATAGCGGGG 300  
DB 38441 GCGGACAGGTGGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGACGATAGCGGGG 38500  
QY 301 GAATACGGCGCGCTGCGACACGAGGC-GATCTTGTTCAGTCTTGTGATTTGGCCGCTCA 359  
DB 38501 GAATACGGCGCGCTGCGACACGAGGCATCTTGTGATTTGGCCGCTCA 38560  
QY 360 CGTGTTCACACGAGTGGTGTGATTCGCGACACGATCAATTCCTTACGACACCGTC 419  
DB 38561 CGTGTTCACACGAGTGGTGTGATTCGCGACACGATCAATTCCTTACGACACCGTC 38620  
QY 420 CAGACCGCGGAGGCGCATACCCGCTTGGCCGCGATAGGGAATGCGCGGTGCACTC 479  
DB 38621 CAGACCGCGGAGGCGCATACCCGCTTGGCCGCGATAGGGAATGCGCGGTGCACTC 38680  
QY 480 GCGGCCAGGTGGCGCTTACCACTGATGCAACCAAGGGGCGCGGTGTGTGATG 539  
DB 38681 GCGGCCAGGTGGCGCTTACCACTGATGCAACCAAGGGGCGCGGTGTGTGATG 38740  
QY 540 GGGGGGTGCGCGCTGCAACCGCGGACGTCGTGTATGCGCGCGGACCGCCGCG 599  
DB 38741 GGGGGGTGCGCGCTGCAACCGCGGACGTCGTGTATGCGCGCGGACCGCCGCG 38800  
QY 600 TACAAGGACCGCGCATGCGCAAGGCGGATGGCGGACCGTTAGGCTTACGATCAAC 659  
DB 38801 TACAAGGACCGCGCATGCGCAAGGCGGATGGCGGACCGTTAGGCTTACGATCAAC 38860  
QY 660 ATGACAACTTGGCACTGCAAGCGGATTCGCGCGGATTCACACATGCTACTCA 719  
DB 38861 ATGACAACTTGGCACTGCAAGCGGATTCGCGCGGATTCACACATGCTACTCA 38920  
QY 720 TGGGCTTAGAGGTGCGGCTCAAAAGTGGCGGACTGGTATTGGGGCGCTGCTG 779  
DB 38921 TGGGCTTAGAGGTGCGGCTCAAAAGTGGCGGACTGGTATTGGGGCGCTGCTG 38980  
QY 780 GGGCGAGGCGCAAGGACCAAAATAGTCTGAATTCCTTCCGATATGAACA 839  
DB 38981 GGGCGAGGCGCAAGGACCAAAATAGTCTGAATTCCTTCCGATATGAACA 39040  
QY 840 GGTGCGGTACTGTGTGATATGACCATGACGAGGCGGTGTTTCAAGGCTCAGACCG 899  
DB 39041 GGTGCGGTACTGTGTGATATGACCATGACGAGGCGGTGTTTCAAGGCTCAGACCG 39100

QY 900 ACCACTACGACACCCGACGCTTGGCTGACGACACGCTTTTACTGCGGTGGCAAC 959  
DB 39101 ACCACTACGACACCCGACGCTTGGCTGACGACACGCTTTTACTGCGGTGGCAAC 39160  
QY 960 ATGCGGCTGGGTGCGGACGACGACGCTACGGGCTGACCAACGGGAGTGGCGAT 1019  
DB 39161 ATGCGGCTGGGTGCGGACGACGACGCTACGGGCTGACCAACGGGAGTGGCGAT 39220  
QY 1020 GTGCTGAGCTTGGCCGACATGGCTGGGCGGCGGCTGCGATTCGGCACTAGCC 1079  
DB 39221 GTGCTGAGCTTGGCCGACATGGCTGGGCGGCGGCTGCGATTCGGCACTAGCC 39280  
QY 1080 AAAGTCTTTGACGACGACGAGGCGCTTACTGTCCGAAGGAGTGGCCACGACTGGG 1139  
DB 39281 AAAGTCTTTGACGACGACGAGGCGCTTACTGTCCGAAGGAGTGGCCACGACTGGG 39340  
QY 1140 GTGCGCTTACGACGCGCGGCGGCTGCTGGGCTGACGCTGGCGGCTGTTAGCGCGAG 1199  
DB 39341 GTGCGCTTACGACGCGCGGCGGCTGCTGGGCTGACGCTGGCGGCTGTTAGCGCGAG 39400  
QY 1200 CACACGTCNGAGTAAGGAACGATGATG 1229  
DB 39401 CACACGTCNGAGTAAGGAACGATGATG 39430

RESULT 12  
A87606  
LOCUS A87606 1235 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 3 from Patent W09836089.  
ACCESSION A87606  
VERSION A87606.1 GI:6736246  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
1 (bases 1 to 1235)  
AUTHORS Flohe, L. and Singh, M.  
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
JOURNAL Patent: W0 9836089-A 20-AUG-1998;  
FLOHE LEOPOLD (DE) SINGH MAHAVIR (DE)  
FEATURES  
Location/Qualifiers  
source 1..1235  
/organism="unidentified"  
/db\_xref="taxon:3264"

BASE COUNT 236 a 394 c 382 g 220 t 3 others  
ORIGIN

Query Match 98.7%; Score 1213.4; DB 5; Length 1235;  
Best Local Similarity 99.6%; Pred. No. 3.4e-165;  
Matches 1225; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 ATCTTGACATTAATGAACTTTCTTACACTGAAAGCTATCGAGGGGTAATC 60  
DB 1 ATCTTGACATTAATGAACTTTCTTACACTGAAAGCTATCGAGGGGTAATC 60  
QY 61 ATGCGGTGGGTATTCGACGAGAACCAAAACAAAGATTCGGGGTGGCCATCAACCCG 120  
DB 61 ATGCGGTGGGTATTCGACGAGAACCAAAACAAAGATTCGGGGTGGCCATCAACCCG 120  
QY 121 GCGGCGTGGGAACTAACCCGTGTGCGCATGAGTGTCTATCGAGGAGTGGCCGA 180  
DB 121 GCGGCGTGGGAACTAACCCGTGTGCGCATGAGTGTCTATCGAGGAGTGGCCGA 180  
QY 181 GAGGCTCGGTATACGACGCGGATTTCAAAGGCGGACGCGCACTGGTGGCAC 240  
DB 181 GAGGCTCGGTATACGACGCGGATTTCAAAGGCGGACGCGCACTGGTGGCAC 240  
QY 241 GCGACACAGGTGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGACGATAGCGGCG 300  
DB 241 GCGACACAGGTGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGACGATAGCGGCG 300

```

OY 301 GAATACGGCCGCTGCGACAGGGG-GATCTGTTCAGCTTCTGCAATTTGGCCGCTCA 359
DB 301 GAAATACGGCCGCTGCGACAGGGG-GATCTGTTCAGCTTCTTTCATTTTGGCCGCTCA 360
OY 360 CGTGTTCAGCAGATGCGCTTGTGATTTCCGACACGCTCAATGGCTTACAGACCGCTC 419
DB 361 CGTGTTCAGCAGATGCGCTTGTGATTTCCGACACGCTCAATGGCTTACAGACCGCTC 420
OY 420 CAGACCGCCGAGGCGGACATACCTCCCTGCTTGGCCGATAGGGAAGTCCCGCTGACTC 479
DB 421 CAGACCGCCGAGGCGGACATACCTCCCTGCTTGGCCGATAGGGAAGTCCCGCTGACTC 480
OY 480 GCGGCCAGCTTGGCGCTTACCACTGATGCAACCAAGGGGCGCGGTGTGTATG 539
DB 481 GCGGCCAGCTTGGCGCTTACCACTGATGCAACCAAGGGGCGCGGTGTGTATG 540
OY 540 GCGGGGCTCCGCGCTGCAACCGGCGAGCTGTGTATGCGGCGCGGCAACCGCGGCG 599
DB 541 GCGGGGCTCCGCGCTGCAACCGGCGAGCTGTGTATGCGGCGCGGCAACCGCGGCG 600
OY 600 TACAACGCAAGCCGCAATGCGCAACGCGGCGGCGGCGGCTTACAGCTCAAC 659
DB 601 TACAACGCAAGCCGCAATGCGCAACGCGGCGGCGGCGGCTTACAGCTCAAC 660
OY 660 ATGCAACAACTTCCGCACTGCAAGCGGCTTCTGCGGCGGCTTCCACACTGCTACTCA 719
DB 661 ATGCAACAACTTCCGCACTGCAAGCGGCTTCTGCGGCGGCTTCCACACTGCTACTCA 720
OY 720 TCGGCTTACGAGCTGAGGGTGGCTTCAAAAGTGGCGGCTGTATTTGGGCGCTCTG 779
DB 721 TCGGCTTACGAGCTGAGGGTGGCTTCAAAAGTGGCGGCTGTATTTGGGCGCTCTG 780
OY 780 GTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 839
DB 781 GTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
OY 840 GGTGCGTACTGCTGATATAGCATATGCAACGAGGCGGCTGTTTCAAGGCTTCAAGCG 899
DB 841 GGTGCGTACTGCTGATATAGCATATGCAACGAGGCGGCTGTTTCAAGGCTTCAAGCG 900
OY 900 ACACCTTAGACACCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 959
DB 901 ACACCTTAGACACCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
OY 960 ATGCGCGCTTCCGCTGCGGAGAGCTGCACTGCGCTGCAACGCGGAGATGCGCTAT 1019
DB 961 ATGCGCGCTTCCGCTGCGGAGAGCTGCACTGCGCTGCAACGCGGAGATGCGCTAT 1020
OY 1020 GTGCTGAGCTTCCGCAACATGCTGCGGCGGCGGCTGCGGCTGCAATCGGCACTAGCC 1079
DB 1021 GTGCTGAGCTTCCGCAACATGCTGCGGCGGCGGCTGCGGCTGCAATCGGCACTAGCC 1080
OY 1080 AAGGCTTTCAGCGATGCAAGAGGGGCGGCTTACCTGCGAAGGGGTGGCGACCGCTAGGG 1139
DB 1081 AAGGCTTTCAGCGATGCAAGAGGGGCGGCTTACCTGCGAAGGGGTGGCGACCGCTAGGG 1140
OY 1140 GTGCGCTTACGCAAGCGCGGCGGCTGCTGCGCTGACTGCTGCGGCGGCTGTTAGCGCGAG 1199
DB 1141 GTGCGCTTACGCAAGCGCGGCGGCTGCTGCGCTGACTGCTGCGGCGGCTGTTAGCGCGAG 1200
OY 1200 CACAGCTCAGAGTAAGGGAAGGATGATG 1229
DB 1201 CACAGCTCAGAGTAAGGGAAGGATGATG 1230

```

## RESULT 13

```

LOCUS A89746 1235 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 3 from Patent WO9832862.
ACCESSION A89746
VERSION A89746.1 GI:6738280
KEYWORDS
SOURCE unidentified.

```

```

ORGANISM unidentified
REFERENCE 1 (bases 1 to 1235)
AUTHORS Flohe, L. and Singh, M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FLOHE LEOBOLD (DE); SINGH MAHAVIR (DE)
FEATURES
Source location/Qualifiers
BASE COUNT 236 a 394 c 382 g 220 t 3 others
ORIGIN
Query Match 98.7% Score 1213.4; DB 5; Length 1235;
Best Local Similarity 99.6% Pred. No. 3.4e-165;
Matches 1225; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
OY 1 ATCTGCAAGATTAATGCACTTCTTCACTGAAAGCTACAGTATGAGAGGGGTATC 60
DB 1 ATCTGCAAGATTAATGCACTTCTTCACTGAAAGCTACAGTATGAGAGGGGTATC 60
OY 61 ATGCGCGTGGTATTCGACGAGACCAAAAAACAAGATTCCGGGTGCCATCACCCCG 120
DB 61 ATGCGCGTGGTATTCGACGAGACCAAAAAACAAGATTCCGGGTGCCATCACCCCG 120
OY 121 GCGGCGCTCCGCGAATCAACCGTGTGCGCATAGGTGCTATCCAGGAGGTGCCGGA 180
DB 121 GCGGCGCTCCGCGAATCAACCGTGTGCGCATAGGTGCTATCCAGGAGGTGCCGGA 180
OY 181 GAGGCGCTGCTATCAGCAGCGGATTTCAAGGCGGAGCGGCGCACTGGTGGGCGC 240
DB 181 GAGGCGCTGCTATCAGCAGCGGATTTCAAGGCGGAGCGGCGCACTGGTGGGCGC 240
OY 241 GCGGACAGGTGTGGGCGGAGCGTGTATTTCTCAAGTCAAAAGAACGATAGCGCG 300
DB 241 GCGGACAGGTGTGGGCGGAGCGTGTATTTCTCAAGGCAAGGCAATAGCGCGCG 300
OY 301 GAATACGGCCGCTGCGACAGGGG-GATCTGTTCAGCTTCTTTCATTTTGGCCGCTCA 359
DB 301 GAATACGGCCGCTGCGACAGGGG-GATCTGTTCAGCTTCTTTCATTTTGGCCGCTCA 360
OY 360 CGTGTTCAGCAGATGCGCTTGTGATTTCCGACACGCTCAATGGCTTACAGACCGCTC 419
DB 361 CGTGTTCAGCAGATGCGCTTGTGATTTCCGACACGCTCAATGGCTTACAGACCGCTC 420
OY 420 CAGACCGCCGAGGCGGCACTACJCTGCTTGGCCGATGAGGGAAGTGGCGGCTGACTC 479
DB 421 CAGACCGCCGAGGCGGCACTACJCTGCTTGGCCGATGAGGGAAGTGGCGGCTGACTC 480
OY 480 GCGGCCAGCTTGGCGCTTACCACTGATGCAACCAAGGGGCGCGGTGTGTATG 539
DB 481 GCGGCCAGCTTGGCGCTTACCACTGATGCAACCAAGGGGCGCGGTGTGTATG 540
OY 540 GCGGGGCTCCGCGCTGCAACCGGCGAGCTGTGTATGCGGCGCGGCAACCGCGGCG 599
DB 541 GCGGGGCTCCGCGCTGCAACCGGCGAGCTGTGTATGCGGCGCGGCAACCGCGGCG 600
OY 600 TACAACGCAAGCCGCAATGCGCAACGCGGCGGCGGCGGCTTACAGCTCAAC 659
DB 601 TACAACGCAAGCCGCAATGCGCAACGCGGCGGCGGCGGCTTACAGCTCAAC 660
OY 660 ATGCAACAACTTCCGCACTGCAAGCGGCTTCTGCGGCGGCTTCCACACTGCTACTCA 719
DB 661 ATGCAACAACTTCCGCACTGCAAGCGGCTTCTGCGGCGGCTTCCACACTGCTACTCA 720
OY 720 TCGGCTTACGAGCTGAGGGTGGCTTCAAAAGTGGCGGCTGTATTTGGGCGCTCTG 779
DB 721 TCGGCTTACGAGCTGAGGGTGGCTTCAAAAGTGGCGGCTGTATTTGGGCGCTCTG 780
OY 780 GTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 839

```



Db 781 GTCCAGAGGCGCAAGGACCCAAATTAGTCTCGAATTCAGTGTGCGGCATATGAACCA 840  
 QY 840 GTGCGGTACTGTGATATATAGCCATGACAGAGGCGGCTGTTTGAAGGCTCAGACCG 899  
 Db 841 GGTGCGGTACTGTGATATATAGCCATGACAGAGGCGGCTGTTTGAAGGCTCAGACCG 900  
 QY 900 AACACTAGACACCCGAGCTTCCGCTGACAGACAGCTGTTTACTGCTGGGGAAC 959  
 Db 901 AACACTAGACACCCGAGCTTCCGCTGACAGACAGCTGTTTACTGCTGGGGAAC 960  
 QY 960 ATGCGCGCTGCTGCGGAGAGAGCTGACCTAGCGCTGACCAAGGAGGATGCCGTAT 1019  
 Db 961 ATGCGCGCTGCTGCGGAGAGAGCTGACCTAGCGCTGACCAAGGAGGATGCCGTAT 1020  
 QY 1020 GTGCTGAGCTTCCGACCATGCTGCGGCGGCGCTGCGGCTGCAATCCGCACTAGCC 1079  
 Db 1021 GTGCTGAGCTTCCGACCATGCTGCGGCGGCGCTGCGGCTGCAATCCGCACTAGCC 1080  
 QY 1080 AAGGCTTTCAGACGACGAAAGGAGGCTTACTGCTCCGAAGGAGGCTGCGCACCTGAGG 1139  
 Db 1081 AAGGCTTTCAGACGACGAAAGGAGGCTTACTGCTCCGAAGGAGGCTGCGCACCTGAGG 1140  
 QY 1140 GTGCGGTTCACGAGCGCGGCGGAGCTGCTGCGCTGACTCTCJGCGGCTGTTAGCGGAG 1199  
 Db 1141 GTGCGGTTCACGAGCGCGGCGGAGCTGCTGCGCTGACTCTCJGCGGCTGTTAGCGGAG 1200  
 QY 1200 CACACGTGCGAGTAAGGAAGCGGATGATG 1229  
 Db 1201 CACACGTGCGAGTAAGGAAGCGGATGATG 1230

RESULT 14  
 A87608  
 LOCUS A87608 1228 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 5 from Patent M09836089.  
 ACCESSION A87608  
 VERSION A87608.1 GI:6736248  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1228)  
 AUTHORS Flohe, L. and Singh, M.  
 TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source 1..1228  
 location/Qualifiers  
 1..1228  
 /organism="unidentified"  
 /db\_xref="taxon:32644"

BASE COUNT 236 a 391 c 382 g 219 t  
 ORIGIN

Query Match 98.6%; Score 1211.9; DB 5; Length 1228;  
 Best Local Similarity 99.7%; Pired. No. 5; 7e-165;  
 Matches 1224; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 ATCTTGAGATTAATCAAGCTTCTTCACTGAGAGGATACAGATATCAAGAGAGGTATATC 80  
 Db 1 ATCTTGAGATTAATCAAGCTTCTTCACTGAGAGGATACAGATATCAAGAGAGGTATATC 60  
 QY 61 ATGCGGCTGCTGATTCGACGAGACCAAAACAAAGATTCGCGGAGGACATACACCCG 120  
 Db 61 ATGCGGCTGCTGATTCGACGAGACCAAAACAAAGATTCGCGGAGGACATACACCCG 120  
 QY 121 GCCGCGCTGCGGAACTTAACCGCTGCTGCGCATGAGGTGCTCATCCAGGAGGTGCCGA 180  
 Db 121 GCCGCGCTGCGGAACTTAACCGCTGCTGCGCATGAGGTGCTCATCCAGGAGGTGCCGA 180  
 QY 181 GAGGCTGCGCTATACCGAGCGCGGATTTCAAGGCGGAGCGCGCAACTGTGCGGACC 240  
 Db 181 GAGGCTGCGCTATACCGAGCGCGGATTTCAAGGCGGAGCGCGCAACTGTGCGGACC 240

QY 241 GCCGACAGGTGTGGCGCGACGCTGATTTATTECTCAAGGTCAAAAGACCGATAGCGCG 300  
 Db 241 GCCGACAGGTGTGGCGCGACGCTGATTTATTECTCAAGGTCAAAAGACCGATAGCGCG 300  
 QY 301 GAATAGGCGCGCTGCGACAGGCGC-GATCTTGTACGTTCTTTCATTTGGCGCGTCA 359  
 Db 301 GAATAGGCGCGCTGCGACAGGCGC-GATCTTGTACGTTCTTTCATTTGGCGCGTCA 360  
 QY 360 CGTGTTCACGATGCTGTTGTGATTCGCGGACACGCAATTCGCTACGAGACCGTC 419  
 Db 361 CGTGTTCACGATGCTGTTGTGATTCGCGGACACGCAATTCGCTACGAGACCGTC 420  
 QY 420 CAGACGCGGAAAGCGCACTACCCCTGCTTGCCTGATGAGCGAAGTCCGCGTCACTC 479  
 Db 421 CAGACGCGGAAAGCGCACTACCCCTGCTTGCCTGATGAGCGAAGTCCGCGTCACTC 480  
 QY 480 GCGCGCGAGGTTGGCGCTTACCACTGATGCGAACCCAAAGGAGGCGCGGCTGCTGATG 539  
 Db 481 GCGCGCGAGGTTGGCGCTTACCACTGATGCGAACCCAAAGGAGGCGCGGCTGCTGATG 540  
 QY 540 GCGCGGCTGCGCGGCTGCGAACCGGCGAGCGTGTGTGATGCGGCGCGGACCGCGGC 599  
 Db 541 GCGCGGCTGCGCGGCTGCGAACCGGCGAGCGTGTGTGATGCGGCGCGGACCGCGGC 600  
 QY 600 TACACGCGACCGCGCATCGCCAAAGGAGGATGCGGCGACCGTTACGCTTACATCAAC 659  
 Db 601 TACACGCGACCGCGCATCGCCAAAGGAGGATGCGGCGACCGTTACGCTTACATCAAC 660  
 QY 660 ATGACAAACTTGGCACTCGACGCGGAGTTGCGGCGCGGATCCACACTGCTACTCA 719  
 Db 661 ATGACAAACTTGGCACTCGACGCGGAGTTGCGGCGCGGATCCACACTGCTACTCA 720  
 QY 720 TCGGCTTACGAGCTGAGGAGTGCCTGCAACGCGCGGACCGTGTGATGAGGCGCGCTG 779  
 Db 721 TCGGCTTACGAGCTGAGGAGTGCCTGCAACGCGCGGACCGTGTGATGAGGCGCGCTG 780  
 QY 780 GTGCGAGGCGCGACCGACCCAAATTAGTCTGCAATTACTGTGCGGATGAAACCA 839  
 Db 781 GTGCGAGGCGCGACCGACCCAAATTAGTCTGCAATTACTGTGCGGATGAAACCA 840  
 QY 840 GTGCGGTACTGTGATATAGCATGACGACGAGGCGGCTGTTTCAAGGCTCAGACCG 899  
 Db 841 GTGCGGTACTGTGATATAGCATGACGACGAGGCGGCTGTTTCAAGGCTCAGACCG 900  
 QY 900 AACACTAGACACCCGAGCTTCCGCTGACAGACAGCTGTTTACTGCGGAGGAGC 959  
 Db 901 AACACTAGACACCCGAGCTTCCGCTGACAGACAGCTGTTTACTGCGGAGGAGC 960  
 QY 960 ATGCGCGCTGCTGCGGAGAGAGCTGACCTAGCGCTGACCAAGCGGAGATGCCGTAT 1019  
 Db 961 ATGCGCGCTGCTGCGGAGAGAGCTGACCTAGCGCTGACCAAGCGGAGATGCCGTAT 1020  
 QY 1020 GTGCTGAGCTTCCGACCATGCTGCGGCGGCGCTGCGGCTGCAATCCGCACTAGCC 1079  
 Db 1021 GTGCTGAGCTTCCGACCATGCTGCGGCGGCGCTGCGGCTGCAATCCGCACTAGCC 1080  
 QY 1080 AAGGCTTTCAGACGACGAAAGGAGGCTTACTGCTCCGAAGGAGGCTGCGCACCTGAGG 1139  
 Db 1081 AAGGCTTTCAGACGACGAAAGGAGGCTTACTGCTCCGAAGGAGGCTGCGCACCTGAGG 1140  
 QY 1140 GTGCGGTTCACGAGCGCGGCGGAGCTGCTGCGCTGACTCTCJGCGGCTGTTAGCGGAG 1199  
 Db 1141 GTGCGGTTCACGAGCGCGGCGGAGCTGCTGCGCTGACTCTCJGCGGCTGTTAGCGGAG 1200  
 QY 1200 CACACGTGCGAGTAAGGAAGCGGATGATG 1227  
 Db 1201 CACACGTGCGAGTAAGGAAGCGGATGATG 1228

RESULT 15  
 A89748  
 LOCUS A89748 1228 bp DNA PAT 22-JAN-2000



DEFINITION Sequence 5 from Patent WO9832862.  
 ACCESSION A89748  
 VERSION A89748.1 GI:6738282  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1228)  
 FLOHE, L. and SINGH, M.  
 TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
 JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source 1..1228  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 236 a 391 c 382 g 219 t  
 ORIGIN

Query Match 98.6%; Score 1211.8; DB 5; Length 1228;  
 Best Local Similarity 99.7%; Pred. No. 5.7e-165;  
 Matches 1224; Conservative 0; Mismatches 4; Indels 1; Gap: 1;

QY 1 ATCTTGACAGTTATGCACTTTCTTACACTGAGCGTACATATCGAGAGGGTAAATC 60  
 DB 1 ATCTTGAGATTATGCACTTTCTTCTATCTAGAGGTACATCGAGAGGGTAAATC 60  
 QY 61 ATCGCGCTCGTATTCGACGCGAGACCAAAAGACGAATTCGGGGTGCATCACCCG 120  
 DB 61 ATCGCGCTCGTATTCGACGCGAGACCAAAAGACGAATTCGGGGTGCATCACCCG 120  
 QY 121 GCCGCGCTCGGAACTAACCCCTCGTGGCCATGAGGTCTCATCCAGCGAGTGGCGGA 180  
 DB 121 GCCGCGCTCGGAACTAACCCCTCGTGGCCATGAGGTCTCATCCAGCGAGTGGCGGA 180  
 QY 181 GAGGGCTCGGCTATACCGGAGCGGGATTTCAGAGCGGACCGCCGCACTGGTGGCACC 240  
 DB 181 GAGGGCTCGGCTATACCGGAGCGGGATTTCAGAGCGGACCGCCGCACTGGTGGCACC 240  
 QY 241 GCCGACGAGGTGCGGCGGACCGCTGATTTATGCTCAAGGTCAAGAACCGATAGCGGC 300  
 DB 241 GCCGACGAGGTGCGGCGGACCGCTGATTTATGCTCAAGGTCAAGAACCGATAGCGGC 300  
 QY 301 GAATPAGGCGCGCTGCGACACGSGC-GATCTGTTCACGTTCTTGCAATTTGGCGCGTCA 359  
 DB 301 GAATPAGGCGCGCTGCGACACGSGC-GATCTGTTCACGTTCTTGCAATTTGGCGCGTCA 359  
 QY 360 CGATGCTGACGAGTGGGTGTTGGATTCGGGACCAAGTGCATTCGCTAGAGACGTC 419  
 DB 361 CGATGCTGACGAGTGGGTGTTGGATTCGGGACCAAGTGCATTCGCTAGAGACGTC 419  
 QY 420 CAGACCGCGGAGCGGACGCTACCTGCTTGGCCGATGAGCGAAGTCCGCGGTGACTC 479  
 DB 421 CAGACCGCGGAGCGGACGCTACCTGCTTGGCCGATGAGCGAAGTCCGCGGTGACTC 479  
 QY 480 GCCGCGGAGTGGCGTTACCACTGATGCGAACCCAGGGGGCGCGGTGTGTGATG 539  
 DB 481 GCCGCGGAGTGGCGTTACCACTGATGCGAACCCAGGGGGCGCGGTGTGTGATG 539  
 QY 540 GGGGGGGTGGCGGCGTGAACCGGGCGAGTGTGTGATTCGGCCCGGACCGCGCGC 599  
 DB 541 GGGGGGGTGGCGGCGTGAACCGGGCGAGTGTGTGATTCGGCCCGGACCGCGCGC 599  
 QY 600 TACAAGCGACCGCGCATCGCAAGGCGATGGCGGACCGTACGGTTAGACATCAAC 659  
 DB 601 TACAAGCGACCGCGCATCGCAAGGCGATGGCGGACCGTACGGTTAGACATCAAC 659  
 QY 660 ATGACAACTTGGCACTGACGCGGAGTTCGCGCGGATCCACACTCGCTACTCA 719  
 DB 661 ATGACAACTTGGCACTGACGCGGAGTTCGCGCGGATCCACACTCGCTACTCA 719  
 QY 720 TGGGCTAGAGCTCGAGGGTGGCGTAAACGTGCCGACCTGCTGATGGGGCGCTCTG 779

DB 721 TGGGCTAGAGCTCGAGGGTGGCGTAAACGTGCCGACCTGCTGATGGGGCGCTCTG 780  
 QY 780 GTGCCAGGCGCCCAAGGACCCCAATAGTCTGGAATTCATCTGTGGCATATGAACCA 839  
 DB 781 GTGCCAGGCGCCCAAGGACCCCAATAGTCTGGAATTCATCTGTGGCATATGAACCA 840  
 QY 840 GTGGCTAGTGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 899  
 DB 841 GTGGCTAGTGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
 QY 900 ACCACCTAGACACCGGACGCTTGGCGGTGACAGACAGCTTTTACTGCTGGCGAAC 959  
 DB 901 ACCACCTAGACACCGGACGCTTGGCGGTGACAGACAGCTTTTACTGCTGGCGAAC 960  
 QY 960 ATGCCCGCTCGTGGCGGAGAGCTGACCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 1019  
 DB 961 ATGCCCGCTCGTGGCGGAGAGCTGACCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 1020  
 QY 1020 GTGCTGAGCTTGGCGGACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1079  
 DB 1021 GTGCTGAGCTTGGCGGACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 QY 1080 AAAGTCTTTCAGACGACGAGAGGGCGTTACTGTCCGAAGGGGTGCCACCGACCTGGGG 1139  
 DB 1081 AAAGTCTTTCAGACGACGAGAGGGCGTTACTGTCCGAAGGGGTGCCACCGACCTGGGG 1140  
 QY 1140 GTGGCTGACGAGCGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1199  
 DB 1141 GTGGCTGACGAGCGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
 QY 1200 CACAGCTGAGAGTAAGGAGACGATGA 1227  
 DB 1201 CACAGCT GAGAGTAAGGAGACGATGA 1228

Search completed: June 22, 2000, 14:55:19  
 Job time: 17734 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:09:11 ; Search time 458.59 Seconds

(without alignments)  
670.503 Million cell updates/sec

Title: US-09-362-485-7

Perfect score: 1229  
Sequence: 1 ATCTTCGAGATTATCGAAC.....GAGTAAGGAGAGCATGATG 1229

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1199.4	97.6	1245	1 V49626	Mycobacterium tube
2	1199.4	97.6	1260	1 V49510	Mycobacterium sp.
3	1199.4	97.6	1260	1 V49625	Mycobacterium tube
4	454	36.9	682	1 V49511	Mycobacterium mari
5	265.2	21.6	1125	1 N91423	Sequence of heat-r
6	203.6	16.6	28171	1 V52155	Streptococcus pneu
7	166.4	13.5	9280	1 V74442	Staphylococcus aur
8	89.2	7.3	1074	1 T67971	H. pylori membrane
9	69.2	5.6	544	1 V43039	Streptococcus pneu
10	56.8	4.6	31096	1 V74370	Streptococcus pneu
11	49.2	4.0	390	1 O21833	Randomising oligon
12	49.2	4.0	390	1 O36859	PCR primer for 5'
13	48.8	4.0	15872	1 T68715	Streptomyces venez
14	46.4	3.8	985	1 V44439	Mycobacterium tube
15	46.4	3.8	985	1 V45448	M. tuberculosis im
16	45.4	3.7	833	1 O64203	snab gene encoding
17	45.4	3.7	5392	1 O64201	Sequence comprisin
18	45	3.7	3946	1 T93610	Mycobacterium tube
19	44.8	3.6	535	1 V44428	Mycobacterium tube
20	44.8	3.6	535	1 V64537	M. tuberculosis im
21	44.4	3.6	2582	1 T73117	Actinoplanes sp. a
22	44.4	3.6	17955	1 V56642	Actinoplanes sp. a
23	43.6	3.5	28598	1 T06769	Sorangium cellulos
24	43.6	3.5	28598	1 T89956	Sorangium cellulos
25	43.6	3.5	49377	1 V05287	The soraphen biosy
26	42.8	3.5	114955	1 X53491	Human adenosine Al
27	42.4	3.4	1833	1 O64206	snbr gene encoding
28	42.4	3.4	2185	1 V84066	Clone p5-4 encodin
29	42.4	3.4	2185	1 O99365	S. lividans protea
30	42.4	3.4	24379	1 T93095	Streptomyces fremo
31	42.4	3.4	24379	1 V25925	Streptomyces roseo
32	42.2	3.4	329	1 V44425	Mycobacterium tube
33	42.2	3.4	329	1 V64334	M. tuberculosis im
34	41.8	3.4	882	1 V44403	Mycobacterium tube

35	41.8	3.4	882	1 V64512	M. tuberculosis im
36	41	3.3	3946	1 T93610	Mycobacterium tube
37	41	3.3	29879	1 O46806	extra region of S.
38	40.8	3.3	2151	1 O76252	HSV-2 protease, IC
39	40.8	3.3	2151	1 O76261	HSV-2 protease/ICP
40	40.8	3.3	2472	1 O84671	HSV-2 UL26 gene. N
41	40.8	3.3	20387	1 V62159	HSV-2 strain SB5 C
42	40.8	3.3	26338	1 V62134	HSV-2 strain SB5 C
43	40.8	3.3	117213	1 V62176	HSV-2 strain SB5 C
44	40.6	3.3	4257	1 V10362	Infected cell prot
45	40.6	3.3	4257	1 V68520	The nucleotide seq

## ALIGNMENTS

RESULT	1
V49626	V49626 standard; DNA; 1245 BP.
AC	V49626;
DT	20-NOV-1998 (first entry)
DE	Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW	ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS	Mycobacterium tuberculosis.
PN	MO9836089-A2.
PD	20-AUG-1998.
PF	29-JAN-1998: E00483.
PA	29-JAN-1997: BP-101338.
PI	(FLOH/) FLOHE L.
PI	Flohe L. Hutter B. Kolk A. Singh M.
DR	WPI: 98-457123/39.
PT	Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT	- useful for, e.g. for diagnosis, differentiation of strains,
PT	monitoring vaccination and identification of mycobacterial
PT	inhibitors
PS	Disclosure; Fig 3.19; 55bp; German.
CC	The Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC	the production of kits for diagnosing tuberculosis (TB) and other
CC	mycobacterial infections in humans or animals. Kits are used for direct
CC	diagnosis of TB on clinical samples (e.g. body fluids) and can
CC	differentiate between pathogenic and non-virulent strains, e.g. for
CC	identifying pathogens of the M. tuberculosis (M. t.) complex. The kit may
CC	also be used to identify substances that inhibit mycobacteria, for
CC	combating epidemics and for vaccination follow-up. Oligonucleotides
CC	derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC	also for culture confirmation of isolated strains and for chromosome
CC	fingerprinting to detect/differentiate between mycobacteria, and for
CC	L-alanine-specific biotransformation reactions. AlaDH is an early
CC	antigen, present extracellularly after only a few days of growth, making
CC	it an ideal drug target.
SO	Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;
Query Match	97.6%; Score 1199.4; DB 1; Length 1245;
Best local Similarity	99.3%; Pred. No. 1.5e+240;
Matches 1227; Conservative	0; Mismatches 2; Indels 7; Gaps 2;
QY	1 ATCTTCGAGATTATCGAAGTTCTTCTACACGTAACCGTACAGTATCGAGGGGTAATC 60
DB	1 ATCTTCGAGATTATCGAAGTTCTTCTACACGTAACCGTACAGTATCGAGGGGTAATC 60
QY	61 ATCTTCGAGATTATCGAAGTTCTTCTACACGTAACCGTACAGTATCGAGGGGTAATC 114
DB	61 ATCTTCGAGATTATCGAAGTTCTTCTACACGTAACCGTACAGTATCGAGGGGTAATC 120
QY	115 ACCCGCGCGCGCGTCCGGAATACCCGCTGCGCATGAGTGTCTATCCAGGAGGT 174
DB	121 ACCCGCGCGCGCGTCCGGAATACCCGCTGCGCATGAGTGTCTATCCAGGAGGT 180
QY	175 GCGGAGAGGGCTCGGCTATACCGAGCGGAGTATTAAGCGCGCAGCGCAACTGGTC 234
DB	181 GCGGAGAGGGCTCGGCTATACCGAGCGGAGTATTAAGCGCGCAGCGCAACTGGTC 240

```

QY 235 GGCACCGCCGACAGGTGGGCGGACGCTGATTATGCTCAAGSTCAAGAACCGATA 294
DB 241 GGCACCGCCGACAGGTGGGCGGACGCTGATTATGCTCAAGSTCAAGAACCGATA 300
QY 295 GCGGGGGAATAGCGCGCTGGGACACGGAGC-GATCTTTACAGCTTCTTGGATTGGCC 353
DB 301 GCGGGGGAATAGCGCGCTGGGACACGGAGATCTTCTTACAGCTTCTTGGATTGGCC 360
QY 354 GCGTACGCTGCTTGCACCGCATGCTTGTGGATTCGGGACACCGATTCGCTACGAG 413
DB 361 GCGTACGCTGCTTGCACCGCATGCTTGTGGATTCGGGACACCGATTCGCTACGAG 420
QY 414 ACCGTCAACACCGCCGAAAGCGCACTACCCCTGCTGCCCGATAGCGAAGTCGCGGT 473
DB 421 ACCGTCAACACCGCCGAAAGCGCACTACCCCTGCTGCCCGATAGCGAAGTCGCGGT 480
QY 474 CGAGTCGCGCCGAGGTGGCGCTTACCACTGATGCGAAACCGAGGGCGCGGTGTG 533
DB 481 CGACTGCGCGCCGAGGTGGCGCTTACCACTGATGCGAAACCGAGGGCGCGGTGTG 540
QY 534 CTGATGGGCGGGGTGCGCGCGTTCGAAACCGCGGACGCTGCTGATTCGGCGCGGAC 593
DB 541 CTGATGGGCGGGGTGCGCGCGTTCGAAACCGCGGACGCTGCTGATTCGGCGCGGAC 600
QY 594 GCGGCTACACGACGACCGCGCATGCGCAACGCGCATGGCGCGACGCTTACGATGAC 653
DB 601 GCGGCTACACGACGACCGCGCATGCGCAACGCGCATGGCGCGACGCTTACGATGAC 660
QY 654 ATCAACATGACAAATTTGGGAACTGAGCGGCGATTCGCGCGGAGTTCACACTGCG 713
DB 661 ATCAACATGACAAATTTGGGAACTGAGCGGCGATTCGCGCGGAGTTCACACTGCG 720
QY 714 TACTCATCGCGCTACGAGCTCGAGGCTGCGTCAACGCTGCGGACCTGCTGATGGGCG 773
DB 721 TACTCATCGCGCTACGAGCTCGAGGCTGCGTCAACGCTGCGGACCTGCTGATGGGCG 780
QY 774 GTCTGCTGCTCAGAGCGCCCAAGGACCCAAATTAATGCTCAATTCCTTGTGGCATATG 833
DB 781 GTCTGCTGCTCAGAGCGCCCAAGGACCCAAATTAATGCTCAATTCCTTGTGGCATATG 840
QY 834 AAACGAGTGGGCTACTGATGATATAGCATGACGAGGCGCGCTTTCGAAGCTCA 893
DB 841 AAACGAGTGGGCTACTGATGATATAGCATGACGAGGCGCGCTTTCGAAGCTCA 900
QY 894 CGACGACCACTACGACACCGGAGCTTCGCGCTGACGACGACGCTGTTTACTGCTG 953
DB 901 CGACGACCACTACGACACCGGAGCTTCGCGCTGACGACGACGCTGTTTACTGCTG 960
QY 954 GCGAACATGCGCGCTCGCTGCGGAGGAGAGCTGACCTACGCGCTACCAACGCGAGATG 1013
DB 961 GCGAACATGCGCGCTCGCTGCGGAGGAGAGCTGACCTACGCGCTACCAACGCGAGATG 1020
QY 1014 CCGTATGCTGAGAGTTCGCGCATGCTGCGGCGGCGGCGGCTGCGATTCGCGCA 1073
DB 1021 CCGTATGCTGAGAGTTCGCGCATGCTGCGGCGGCGGCGGCTGCGATTCGCGCA 1080
QY 1074 CTAGCCAAAGGCTTTCGACGACGAGAGGCGCTTACTGTCGAAAGGCTGGCCACGAC 1133
DB 1081 CTAGCCAAAGGCTTTCGACGACGAGAGGCGCTTACTGTCGAAAGGCTGGCCACGAC 1140
QY 1134 CTGGGGGTGCGCTTACGACGACGCGCGGCTGCTGCTGCTGCTGCTGCTTAC 1193
DB 1141 CTGGGGGTGCGCTTACGACGACGCGCGGCTGCTGCTGCTGCTGCTGCTTAC 1200
QY 1194 GCGGAGCACAGCTGCGGAGTAAAGGGAAGGATGATG 1229
DB 1201 GCGGAGCACAGCTGCGGAGTAAAGGGAAGGATGATG 1236

```

RESULT 2  
 ID V49510 standard; DNA: 1260 BP.  
 AC V49510;

```

DE 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlADH DNH.
KM Alanine dehydrogenase; AlADH; ADH; diagnosis; tuberculosis; pathogen;
KM swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN M09832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH) FLOHE L.
PA Flohe L, Hutter B, Kolk A, Singh M;
DR WPI: 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure: Page 11: 57P; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic fliJ sequences. ADH is an antigen that is secreted
CC early during infection
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 97.6%; Score 1199.4; DB 1; Length 1260;
Best Local Similarity 99.3%; Pred. No. 1.5e-240;
Matches 1227; Conservative 0; Mismatches 2; Indels 7; Gaps 2;

QY 1 ATCTGCGATTAATCGAATCTTCTTCACTGAAAGCTATAGTATCGAGAGGGTATC 60
DB 16 ATCTGCGATTAATCGAATCTTCTTCACTGAAAGCTATAGTATCGAGAGGGTATC 75
QY 61 ATGCGCTGCGATATCCAGCGAGACCAAAACAAAG-----AATTCGGGTGGCCATC 114
DB 76 ATGCGCTGCGATATCCAGCGAGACCAAAACAAAGATTCGAATTCGGGTGGCCATC 135
QY 115 ACCCGCGCGCGCTGCGGAACTAACCCGCTGCGCATGAGAGTCTCATCAGGACGT 174
DB 136 ACCCGCGCGCGCTGCGGAACTAACCCGCTGCGCATGAGAGTCTCATCAGGACGT 195
QY 175 GCGGAGAGGGCTGCGCATATCAGCGAGCGGATTTCAAGGCGGCGGCAACTGCTC 234
DB 196 GCGGAGAGGGCTGCGCATATCAGCGAGCGGATTTCAAGGCGGCGGCAACTGCTC 255
QY 235 GGCACCGCGGACAGAGTGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGAACCGATA 294
DB 256 GGCACCGCGGACAGAGTGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGAACCGATA 315
QY 295 GCGGCGGAATACGCGCGCTCGACACGCGGCG-GATCTTTGTTACAGTTCTTCATTTGGCC 353
DB 256 GGCACCGCGGACAGAGTGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGAACCGATA 315
QY 316 GCGGCGGAATACGCGCGCTCGACACGCGGAGATCTTGTACAGTTCTTCATTTGGCC 375
DB 316 GCGGCGGAATACGCGCGCTCGACACGCGGAGATCTTGTACAGTTCTTCATTTGGCC 375
QY 334 GCGTACAGCTTGCACGAGTGGCTTGTGGATTCGGGACACGAGTCAATGCTTAGAG 413
DB 376 GCGTACAGCTTGCACGAGTGGCTTGTGGATTCGGGACACGAGTCAATGCTTAGAG 435
QY 414 ACCGTCAACACCGCCGAAAGCGCACTACCCCTGCTGCCCGATAGCGAAGTCGCGGT 473
DB 436 ACCGTCAACACCGCCGAAAGCGCACTACCCCTGCTGCCCGATAGCGAAGTCGCGGT 495
QY 474 CGACTGCGCGCGCGAGTGGCGCTTACCACTGATGCGAAACCAAGGGGCGGCGGTG 533
DB 496 CGACTGCGCGCGCGAGTGGCGCTTACCACTGATGCGAAACCAAGGGGCGGCGGTG 555
QY 534 CTGATGGGCGGGGTGCGCGCGTTCGAAACCGCGGACGCTGCTGATGGGCGGCGGAC 593
DB 556 CTGATGGGCGGGGTGCGCGCGTTCGAAACCGCGGACGCTGCTGATGGGCGGCGGAC 615

```

QY	594	GCCGGCTACAAAGCAGGCCGCGCATCCGCAACGGCATGGGGGCGACGGTTACGTTCTTAGAC	653
Db	616	GCCGGCTACAAAGCAGGCCGCGCATCCGCAACGGCATGGGGGCGACGGTTCTTAGAC	675
QY	654	ATCAACATGACAAACCTTGGGCACTCGAACGCCAGTTCGCGGCGGATCCACACTCGC	713
Db	676	ATCAACATGACAAACCTTGGGCACTCGAACGCCAGTTCGCGGCGGATCCACACTCGC	725
QY	714	TACTCATCGGCGCTAGAGAGTCCGAGGGTGGCCGTAAACGTCGCCGCACTGGTGATTGGGGCC	773
Db	736	TACTCATCGGCGCTAGAGAGTCCGAGGGTGGCCGTAAACGTCGCCGCACTGGTGATTGGGGCC	795
QY	774	GTCCTGTGTCCAGGCGCCCAAGCGACCCAAATTAATGTCTCGAATTCACCTTGCGCCGCAATG	833
Db	796	GTCCTGTGTCCAGGCGCCCAAGCGACCCAAATTAATGTCTCGAATTCACCTTGCGCCGCAATG	855
QY	834	AAACCAAGTGGCGGTACTGTGTATATAGCCATGACCCAGGCGCGGTCTTTCGAAAGCTCA	893
Db	856	AAACCAAGTGGCGGTACTGTGTATATAGCCATGACCCAGGCGCGGTCTTTCGAAAGCTCA	915
QY	894	CGACGACCACTACGAGACACCCGAGCTTCGCCGTGCAGACACGGTGTATTACTGGGTG	953
Db	916	CGACGACCACTACGAGACACCCGAGCTTCGCCGTGCAGACACGGTGTATTACTGGGTG	975
QY	954	GCGAACATGCGCGGCTCGGTGCGCGGAAGAGCGTGCACCTAGCGCTGACCAACGCCAGCATG	1013
Db	976	GCGAACATGCGCGGCTCGGTGCGCGGAAGAGCGTGCACCTAGCGCTGACCAACGCCAGCATG	1035
QY	1014	CCGTATGTCTGTGAGAGTTCGCCACCACTGGGCTGGCGGGGCGCGGTGCGATCCGGCA	1073
Db	1036	CCGTATGTCTGTGAGAGTTCGCCACCACTGGGCTGGCGGGGCGCGGTGCGATCCGGCA	1095
QY	1074	CTAGCCAAAGGTCTTTTCGACGACGACGAAGGGGCGCTTACTGTCCGACAGGGGTGGCC	1133
Db	1096	CTAGCCAAAGGTCTTTTCGACGACGACGAAGGGGCGCTTACTGTCCGACAGGGGTGGCC	1155
QY	1134	CTGGGGGTCGCCGTTCAACGACGCCGCCGACAGGTGCTGTGGCTGACTCTGGGCGGCTCGTTAC	1193
Db	1156	CTGGGGGTCGCCGTTCAACGACGCCGCCGACAGGTGCTGTGGCTGACTCTGGGCGGCTCGTTAC	1215
QY	1194	GCCGAGCACACGTCNMGAGTAAGGGGAAGGATGATG	1229
Db	1216	GCCGAGCACACGTCGCGGACTAAGGAGGAGATGATG	1251
RESULT	3		
ID	V49625		
ID	V49625	standard; DNA; 1260 BP.	
AC	V49625		
DT	20-NOV-1998	(first entry)	
DE	Mycobacterium tuberculosis	Alanine dehydrogenase.	
KW	ss; Alanine dehydrogenase; tuberculosis; TB; Mycobacteria; L-alanine.		
OS	Mycobacterium tuberculosis.		
PN	MO9836089-A2.		
PD	20-AUG-1998.		
PE	29-JAN-1998.	E00483.	
PR	29-JAN-1997.	EP-101338.	
RA	(FLOH/) FLOHE L.		
FI	Flohe L, Hutter B, Kolk A, Singh M;		
DR	WPI; 98-457123/39.		
PT	Diagnosis of, e.g., tuberculosis from alanine dehydrogenase activity		
PT	- useful for, e.g., for diagnosis, differentiation of strains,		
PT	monitoring vaccination and identification of mycobacterial		
PT	inhibitors		
PS	Clam 13; Fig 2.3; 55pp. German.		
CC	The Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in		
CC	the production of kits for diagnosing tuberculosis (TB) and other		
CC	mycobacterial infections in humans or animals. Kits are used for direct		
CC	diagnosis of TB on clinical samples (e.g. body fluids) and can		
CC	differentiate between pathogenic and non-virulent strains, e.g. for		
CC	identifying pathogens of the M. tuberculosis (M. t.) complex. The kit may		
CC	also be used to identify substances that inhibit Mycobacteria, for		
CC	combating epidemics and for vaccination follow-up. Oligonucleotides		

CC derived from AlADH are used similarly in diagnostic hybridisation tests,  
CC also for culture confirmation of isolated strains and for chromosome  
CC fingerprinting to detect/differentiate between mycoplasmas, and for  
CC L-alanine-specific biotransformation reactions. AlADH is an early  
CC antigen, present extracellularly after only a few days of growth, making  
CC it an ideal drug target.

CC Sequence	1260 BP:	243 A:	403 C:	389 G:	225 T:
-------------	----------	--------	--------	--------	--------

Query Match	97.6%;	Score 1199.4;	DB 1;	Length 1260;
Best Local Similarity	99.3%;	Pred. No. 1.5e-240;		
Matches 1227;	Conservative	0;	Mismatches 2;	Indels 7; Gaps 2;

QY	1	ATCTTGCAAGATTAAACGAACTTTCTTCACACTGAAGCGCTACGATATCGAGAGGGGTAATC	60
Db	16	ATCTTGCAAGATTAAACGAACTTTCTTCACACTGAAGCGCTACGATATCGAGAGGGGTAATC	75
QY	61	ATGCGCGTGGATTATCCACGAGACCCAAAAACAACG-----AATTCGAGGTGGCCATC	114
Db	76	ATGCGCGTGGATTATCCACGAGACCCAAAAACAACGAAATTCATATCCGGGTGGCCATC	135
QY	115	ACCCCGCGCGCGCTGCGGGAACTAACCGGTGTGGCCATGAGTGTCTCATCCAGCGAGT	174
Db	136	ACCCCGCGCGCGCTGCGGGAACTAACCCGCTGTGGCCATGAGTGTCTCATCCAGCGAGT	195
QY	175	GCCGGAGAGGGCTGGGTATCCACCACCGGGATTTCAAGGGGGAGGGGGGCACTGGTC	234
Db	196	GCCGGAGAGGGCTGGGTATCCACCACCGGGATTTCAAGGGGGAGGGGGGCACTGGTC	255
QY	235	GGCACCGCCGACCAAGGTGTGGCCGACGCTGATTTATTGCTCAAGTCAAAAGCCGATA	294
Db	256	GGCACCGCCGACCAAGGTGTGGCCGACGCTGATTTATTGCTCAAGTCAAAAGCCGATA	315
QY	295	GCGGCGAATAGCGCGCGCTGCGAACCGGCG-GATCTTGTATACGTTCTTGATTTGGCC	353
Db	316	GCGGCGAATAGCGCGCGCTGCGAACCGGCGAGATCTTGTACGTTCTTGATTTGGCC	375
QY	354	GCGTACCGTGTTCGACCGATGCGTTGTTGGATTCGGGACCGCATTTGCTTACGAG	413
Db	376	GCGTACCGTGTTCGACCGATGCGTTGTTGGATTCGGGACCGCATTTGCTTACGAG	435
QY	414	ACCGTCCAGACCGCGCGAAGCGCACTACCCCTGCTTGGCCCCGATGAGCGAACTGCCGT	473
Db	436	ACCGTCCAGACCGCGCGAAGCGCACTACCCCTGCTTGGCCCCGATGAGCGAACTGCCGT	495
QY	474	CGACTCGCGCGCGAGTGGGCGCTTACCAACCTGATTCGCAACCCAGAGCGGCGCGGTGTG	533
Db	496	CGACTCGCGCGCGAGTGGGCGCTTACCAACCTGATTCGCAACCCAGAGCGGCGCGGTGTG	555
QY	534	CTGATGGCGGGGGTCCCAACCTGTCACACCAACGTCGTGTGATCGGCGCCGGACCC	593
Db	556	CTGATGGCGGGGGTCCCAACCTGTCACACCAACGTCGTGTGATCGGCGCCGGACCC	615
QY	594	GCGGCGTCAACGCAAGCCCGCATTCGCCCAACGCGATGGGCGCACCGTTACGTTACAGC	653
Db	616	GCGGCGTCAACGCAAGCCCGCATTCGCCCAACGCGATGGGCGCACCGTTACGTTACAGC	675
QY	654	ATCAACATCGACAACTTCGAGCACTCGACGCGAGATTGTGCGGCGGATTCGACATGAG	713
Db	676	ATCAACATCGACAACTTCGAGCACTCGACGCGGAGATTGTGCGGCGGATTCGACATGAG	735
QY	714	TACATATGGGCTACGAGCTGAGGGTGGCGGTCAAAAGCGCGACCTGGTATGGGGGC	773
Db	736	TACATATGGGCTACGAGCTGAGGGTGGCGGTCAAAAGCGCGACCTGGTATGGGGGC	795
QY	774	GTCCTGTGTCAGGCGCCCAAGGCAACCAATTAAGTCTCGAATTCATTGTGCGCGATAG	833
Db	796	GTCCTGTGTCAGGCGCCCAAGGCAACCAATTAAGTCTCGAATTCATTGTGCGCGATAG	855
QY	834	AAACCGAGTGGCGTACTGTGTGATTAAGCATTCGACACGAGGGGCGTGTTCGAAGGCTTA	893
Db	856	AAACCGAGTGGCGTACTGTGTGATTAAGCATTCGACACGAGGGGCGTGTTCGAAGGCTTA	915

QY 894 CGACCGACACCTACGACACCGGACCTTCGCGTGCACGACACGCTGTTTACTGCGTG 953  
 |||||  
 Db 916 CGACCGACACCTACGACACCGGACCTTCGCGTGCACGACACGCTGTTTACTGCGTG 975  
 |||||  
 QY 954 GCGAACATGCGCGCTTCGCGGAGAGCTGCACTACCGCTGACCAAGCGAGATG 1013  
 |||||  
 Db 976 GCGAACATGCGCGCTTCGCGGAGAGCTGCACTACCGCTGACCAAGCGAGATG 1035  
 |||||  
 QY 1014 CCGATATGCTGCGAGCTGCGACATGCTGCGGCGGCTGCGGCTGGAATCCGCGA 1073  
 |||||  
 Db 1036 CCGATATGCTGCGAGCTGCGACATGCTGCGGCGGCTGCGGCTGGAATCCGCGA 1095  
 |||||  
 QY 1074 CTAGCCAAAGGCTTTTCGACGACGAAAGGCGCTTACTGTCGAAAGGCTGCGCCACCGAC 1133  
 |||||  
 Db 1096 CTAGCCAAAGGCTTTTCGACGACGAAAGGCGCTTACTGTCGAAAGGCTGCGCCACCGAC 1155  
 |||||  
 QY 1134 CTGGGGGTGCGCTTCACGCGGCGCGAGCGTGTGCGCTGACTCTGCGCGCTGTTAC 1193  
 |||||  
 Db 1156 CTGGGGGTGCGCTTCACGCGGCGCGAGCGTGTGCGCTGACTCTGCGCGCTGTTAC 1215  
 |||||  
 QY 1194 GCGGACGACAGTCGCGAGTAAAGGAGCGATGATG 1229  
 |||||  
 Db 1216 GCGGACGACAGTCGCGAGTAAAGGAGCGATGATG 1251  
 |||||

RESULT 4  
 ID V49511 standard; DNA; 682 BP.  
 AC V49511:  
 DT 20-OCT-1998 (first entry)  
 DE Mycobacterium marinum Mar3 DNA.  
 KW Alantine dehydrogenase; ADH; ADH; diagnosis; tuberculosis; pathogen;  
 KM swimmers disease; vaccine; epidemic; infection; identification; ss.  
 OS Mycobacterium marinum.  
 FH Key Location/Qualifiers  
 FT 1.682  
 FT /tag- a  
 FT /codon\_start- 3  
 FT /product- "Mar3"  
 FT /note- "Alantine dehydrogenase"  
 PN W09832862-AZ.  
 PD 30-JUL-1998  
 PF 29-JAN-1998; E00484.  
 PR 29-JAN-1997; EP-101339.  
 PA (FLOH/) FLOHE L.  
 PI Flohe L, Hutter B, Kolik A, Singh M;  
 DR MPI; 98-427958/36.  
 DR P-PSDB; W64481.  
 PT Nucleic acid encoding alantine dehydrogenase of Mycobacterium marinum  
 PT used for diagnosis of tuberculosis and other mycobacterial  
 PT diseases, also for treatment and prevention, for drug screening and  
 PT for bio-transformation.  
 PS Claim 1: Page 34-35; 57pp; German.  
 CC This sequence encodes an alantine dehydrogenase (ADH) protein, Mar3  
 CC isolated from Mycobacterium marinum. This protein is used to diagnose  
 CC tuberculosis and other mycobacterial infections (including 'swimmers'  
 CC disease', caused by M. marinum, a fish pathogen) in humans or animals.  
 CC The protein can also be used for control of epidemics and for  
 CC vaccination, to screen for agents with anti-mycobacterial activity, and  
 CC in bio-transformations that are specific for L-alanine. Also mycobacteria  
 CC can be identified by analysis of genomic ADH sequences. ADH is an antigen  
 CC that is secreted early during infection.  
 CC Sequence 682 BP; 105 A; 254 C; 225 G; 98 T;

Query Match 36.9%; Score 454; DB 1; Length 682;  
 Best Local Similarity 80.1%; Pred. No. 4.6e-86;  
 Matches 546; Conservative 0; Mismatches 135; Indels 1; Gaps 1;

QY 98 AATTCGCGTGGCATTACCCCGCGCTCGCGAATACCCGCTGTCGCTATGAG 157  
 |||||  
 Db 1 AATTCGCGTGGCATTACCCCGCGCTCGCGAATACCCGCTGTCGCTATGAG 60  
 |||||

QY 158 TGCTATCCAGGAGGTGCGGAGAGGCTCGGCTATCAGACGCGGATTTCAAGGGCG 217  
 |||||  
 Db 61 TGCTATCCAGGAGGTGCGGAGAGGCTCGGCTATCAGACGCGGATTTCAAGGGCG 120  
 |||||  
 QY 218 CAGGCGGCACTGATGTCGCGACCGCGGACGAGTGTGGGCGGACGCTGATTTATGCTCA 277  
 |||||  
 Db 121 CCGGTGCCAGCTGATCAGACCGCGGACGAGTGTGGGCGGACGCTGATTTATGCTCA 180  
 |||||  
 QY 278 AGGTCAAAACCGATAGCGCGGGAATACGCGGCTGCGACAGGGC- GATCTGTTCA 336  
 |||||  
 Db 181 AGGTCAAAACCGATAGCGCGGGAATACGCGGCTGCGACAGGGC- GATCTGTTCA 240  
 |||||  
 QY 337 CGTTCTTGATTTGGCGGCTGACGCTGCTGACCGATGCTGTTGATTCGGGACCA 396  
 |||||  
 Db 241 CCGTCTTGATTTGGCGGCTGACGCTGCTGACCGATGCTGTTGATTCGGGACCA 300  
 |||||  
 QY 397 CGTCAATTCCTTACGAGACGCTGCGACGCGGAGGCGGACCTACCCCTGTCGCGGA 456  
 |||||  
 Db 301 CGTCAATTCCTTACGAGACGCTGCGACGCGGAGGCGGACCTACCCCTGTCGCGGA 360  
 |||||  
 QY 457 TGAGCAAGTCCCGGCTGCGACTGCGCGGCGGAGGCTGCGGCTTACCACTGATGCAAGCC 516  
 |||||  
 Db 361 TGAGCAAGTCCCGGCTGCGACTGCGCGGCGGAGGCTGCGGCTTACCACTGATGCAAGCC 420  
 |||||  
 QY 517 AAGGGGCGCGGCTGCTGATGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGG 576  
 |||||  
 Db 421 AAGGGGCGCGGCTGCTGATGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGG 480  
 |||||  
 QY 577 TGATCGGCGCGGCGGCTGCTGATGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGG 636  
 |||||  
 Db 481 TGATCGGCGCGGCGGCTGCTGATGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGG 540  
 |||||  
 QY 637 CGGTTAGCTTACGACATGACATGACATGACATGACATGACATGACATGACATGACATG 696  
 |||||  
 Db 541 TGCTACCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
 |||||  
 QY 697 GCGGATCCACACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 756  
 |||||  
 Db 601 GCGGATCCGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
 |||||  
 QY 757 ACCTGCTGATTTGGGCGGCTGCT 778  
 |||||  
 Db 661 ACATGCTGATTTGGGCGGCTGCT 682  
 |||||

RESULT 5  
 ID N91423 standard; DNA; 1125 BP.  
 AC N91423:  
 DT 01-FEB-1991 (first entry)  
 DE Sequence of heat-resistant alantine dehydrogenase (AH) gene with mol. wt.  
 DE 2 MD or less  
 KW Enzyme; ds.  
 OS Bacillus stearothermophilus IFO 12550.  
 PN J01043194-A.  
 PD 15-FEB-1989.  
 PF 10-AUG-1987; 200524.  
 PR 10-AUG-1987; JP-200524.  
 PA (NIRA) Unilika KK.  
 DR MPI: 89-096095/13.  
 PT Recombinant plasmid for transforming Escherichia coli -  
 PT obtd. by connecting heat resistant alantine dehydrogenase gene to  
 PT vector plasmid  
 PS Disclosure: Fig 2, p693; 9pp; Japanese.  
 CC A recombinant plasmid contg. heat-resistant alantine dehydrogenase (AH)  
 CC gene with mol. wt. 2 or less MD is claimed. Cells transformed with the  
 CC vector produce high levels of heat-resistant AH. Transformed E. coli  
 CC is useful as a clinical inspection reagent.  
 CC Sequence 1125 BP; 246 A; 295 C; 367 G; 217 T;

Query Match 21.6%; Score 265.2; DB 1; Length 1125;  
 Best Local Similarity 56.1%; Pred. No. 5.5e-47;





QY	360	CGTGGTTTCACACCGATGGGTTGTGGATTCGCGGACACCGTCATTTCCATACGAGACCGTC	419
Db	21414	CCAGATTATAGCAAGATGCGTATGTT-----TACACCAAAAAACAACTGAAGTGT	21367
QY	420	CAGACCCCGCAAGCGCACACTACCCCTGCTTGCOCGATGAGCGAAGTCGCCGCTGACTC	479
Db	21366	CGTGACCAATCAAGGACAACATCCGCTCTCTGTTCTCTATGATAGGTATGCGAGTGTATG	21307
QY	480	GCCGCCCGAGTTGGCGGCTTACCATCTGATGGAACCCAAAGGGGGCGGGTGGCTGATG	539
Db	21306	GCTGTTCAATTGCGAAGCTCACTTCTTACTACTAGCAAGCTGTGTGGCTGTGTTTACTT	21247
QY	540	GGCGGGGTGCGCGGCGTGCAGACCGGCGGAGTGTGTGTATTCGCGCGGACACCGCGG	599
Db	21246	GGTGGTGTACCAAGGTGTTCCAAAAGAAAAGTAACTATCATCGGTGTGTGTGCGCGGT	21187
QY	600	TACAAAGAGCGCGGATTCGCCCAAGGAGATGGCGCGACCGCTTACGGTTCTAGACATCAAC	659
Db	21186	ACACATGTGTGCGCGCATGCGCTTGTGCTTGTGTCTCAAGTGTACTATTTTATGATTA	21127
QY	660	ATCGACAAATCTGCGCACTGACGCCGAGTTCGCGCGGATCCACACTGCGTACTCA	719
Db	21126	TCCAAAGCGTCTCTCAGTTCTGAGAAGAGTCTTTGGAAGTCAAAATTCAAACTGTATGCT	21067
QY	720	TGCGCTACGAGCTCGAGGGTGGCGTCAAAAGTCCGCACTGGTATTTGGGGCCGCTCTG	779
Db	21066	AATTCATTCACATTTGAAGCAAGTGTAGAGATGCTGATGTTGGTATGTGAGCCATCTTC	21007
QY	780	GTGCGAGCGCGCAAGGACACCCAAATTTAGTCTCGAATTCACLTCCCGATATGAACCA	839
Db	21006	ATCCGTGGTGAAGAAAGCAGCGGATTTGTGACAGATAGATGATCAAAACAAATGCGTCA	20947
QY	840	GGTGGGCTACTGTGTGATATACCATGACCCAGGGCGGCTGTTGCA---AGGCTACGA	896
Db	20946	GGCTCTGTA-TCGTTGACGTTGCTGTGACCAAGTGGCCGTTTCCAAAACAGTGCACCT	20888
QY	897	CGGACACACTACGACACACCCGAGCTTGCGCGGTGACACACGCTGTTTACTCGTGGCG	956
Db	20887	GTGACACGCAAGATGAACCCGTTATATGAAAAACGSGTTCCTCCACTATGCGGTTGCC	20828
QY	957	AACATGCCCCGCTGTGTCGCGAGACGTGACCTACGCGCTGACGACGCGAGATGCCG	1016
Db	20827	AATATCCCTGGTGGGTTGCTGCGCACTTCAACATCCGCTTACCAATATCTCATCTCT	20768
QY	1017	TATGTGCTCGAGCTTGCGGACCACTGGT	1044
Db	20767	TATATCAAGCTTTGGCTGGCCAAAGAT	20740
RESULT	7		
ID	V74442/c		
AC	V74442: standard: DNA: 9280 BP.		
DE	16-MAR-1999 (first entry)		
RE	Staphylococcus aureus confliq SEQ ID #131.		
KW	Computer Readable medium; vaccine; S.aureus infection; immunodetection;		
KW	cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;		
KW	skin infection; surgical wound infection; scaled skin syndrome;		
KW	toxic shock syndrome; ds.		
OS	Staphylococcus aureus.		
FT	Key	Location/Qualifiers	
FT	misc.feature	841..900	
FT		/*lag- a	
FT		/*lag- b	
FT		/*lag- c	
FT	misc.feature	2641..2700	
FT		/*lag- b	
FT		/*lag- c	
FT		/*lag- d	



Db	9043	AAAGTAATTACTANTTGCATATAGAGACTGTCAGTATACAGACGAGTACTTACCATTTGTA	898
Qy	450	GCCCGGATGACGAAATGCCGGGTCACACTGCCGCCAGAGTTAGCGCTTACCACTGATG	509
Db	8983	TCACCAAAAGATGTAGGTAGCAGGAAAGATGTCAAGTTAGGCGCAGATTCCTACAA	8922
Qy	510	CGAAGCCCAAGGGGACCCGCGTGTCTATGTGGGGGGGTGCGCGCGCTCGAACCGGCGAC	569
Db	8923	AAACTTAATGGTGTATAGGGAATTTCTACTAGTGTGTGTCCAGAGATACCTTAAGGTTAA	8866
Qy	570	GTCGTGTGATGCGCGCCGCGACCCGCGCTACACAGCAGACCCCGATCGCCAAAGGCAATG	629
Db	8863	GTAAGTATATATCGGTGTGGTCAAGCAGGAAACAAAGTCAAGCTAAATTTCACTAGAGCTA	8804
Qy	630	GGCGGACCGCTTAGCGTTCTAGATCAATCAATGACAAACTTCGCAACTGCAGCCGACG	689
Db	8803	GGTGAGATGTTCAGATTTTATGATTTATCCAAAGCTTTACACAAATTAATATATATTTA	8744
Qy	690	TTCTGCGCGCCGAGATCCACACTGCTACTCATCGCGCTAGAGGCTCGAGGGTCCGCTCAA	749
Db	8743	TTTCGTGAGCTGTACATACATTAATATGTCAAAATCCGTGAAATTTGATGTATGTAA	868
Qy	750	CGTGCCGACCTGTGTGATTTGGGGCCGTCTGTGTCCAGGCGCCAAAGCAGCCAAATTAGTC	809
Db	8683	CAAACTGATTTAGTAAATAGGTGACGTTTATTTTCACAGGTCTAAACCGCCAAAGACTTGA	8622
Qy	810	TCGAATTCATCTGTGCGGCATATGAACAGAGTGGCGTACTGGTGTATTTATCCATCGAC	869
Db	8633	ACAGAGACATGATTTAAACAAATGAAAATATGGTCAAGTTATTTATACATTCTATGTAT	8566
Qy	870	CAGGCGCGCTGTTTCGAAGGCTC---ACGACGACACACATAGACACACCCGACGTTGCGC	926
Db	8563	CAAGCGGATATTTTGTGAACACATGATTAATATAGCACATGATGATCTCTACATATAT	8504
Qy	927	GTCGACGACAGCGCTTTTACTAGTGTGGCAATATCCCGCCCTCGGTGCCGAAGACGTG	986
Db	8503	AAGCATGCTGTGGTCAATTATGACAGTTGCAAAATATCCAGGTGCACTACCGGCTACTCG	8444
Qy	987	ACCTACGCGCTGACCAACGCGACGATGCCATATGTGCTGC AGCTTCCGACCATGGCTGG	1048
Db	8443	ACGTTAGCTTAATTAATATGTCAGCCATACCTTATGCGCTATAGCTAATTAAGGATAT	8388
Qy	1047	CGGCGCGCTGCGCGCTGCAATCCGGGCACTAGCCAAAGTCTTTGACGACGACGAGG	1102
Db	8383	AGAGACGATTTAATCAATCAATCAACCATTTATCTATTAGTTAAATCTTACAAAGG	8328
RESULT	8		
Id	T67971	standard; DNA; 1074 BP.	
AC	T67971;		
DT	15-JUL-1997	(first entry)	
DE	H. pylori membrane protein ORF 05CP20518orf61.		
KW	Vaccine; prevention; treatment; infection; identification; binding compound; bacterium; life cycle; activator; bactericidal inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; membrane; amino acid; metabolism; ds.		
CV	member		
CH	Helicobacter pylori.		
PH	Key	Location/Qualifiers	
FT	cds	1..1074	
FT		/*tag= a	
FT		/note= "no stop codon given"	
PN	MO9640893-A1.		
PD	19-DEC-1996		
PF	06-JUN-1996; U09122.		
PR	07-JUN-1995; US-487032.		
PR	01-APR-1996; US-630405.		
PA	(ASTR ) ASTRA AB.		
PI	Berglundh OT, Smith D, Mellgaerd BL;		
DR	Wpi: 97-052306/05.		
DR	P-PSDB: W20718.		
PT	Helicobacter pylori nucleic acid sequences and related polypeptide(s) - useful for vaccines to treat or prevent H. pylori		

PT Infection and to detect Helicobacter  
PS Claim 27, Page 819, 1481bp. English.  
CC The present sequence encodes a Helicobacter pylori membrane  
CC protein likely to contain four membrane spanning regions.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The genomic sequence of H. pylori (ATCC 55579) was determined from  
CC overlapping contigs generated by mechanically shearing the  
CC bacterial DNA. The sequences were analysed for ORF of at least 180  
CC nucleotides, and the predicted coding regions defined by computer  
CC evaluation. To identify likely H. pylori antigens for vaccine  
CC development, the amino acid sequences predicted from various ORF  
CC were analysed for significant homology to other known or exported  
CC membrane proteins. Having identified and determined the sequences  
CC of interest, particular regions can be isolated from H. pylori by  
CC PCR amplification for recombinant polypeptide production, e.g. in  
CC E. coli hosts.  
SQ Sequence (1074 BP): 302 A; 189 C; 297 G; 286 T;

Query Match	7.38;	Score 89.2;	DB 1;	Length 1074;
Best Local Similarity	46.58;	Pred. No. 2e-10;		
Matches 373;	Conservative 0;	Mismatches 413;	Indels 16;	Gaps 2;

OY	265	GATTATTGCTCAAGGTCGAAGAAGCCGATACGGGCGGATATCGGCGCGCTGGCGA - CACGG	332
Db	133	GATTTGGTGCTCAATGCANAGACCTTTTAGCATGATATACCTTTGGCTCAAGAAAA	132
OY	324	GCAGTCCTGTTACAGTCTCTTGCCATTTTGGCCGCGTCACTGCTGGACACGATGGCTTTG	383
Db	193	GCAGCTCTGTTAGTTATTTTGATTTAGCGATCAAAAAACCTTGCTGAATGTTATT	252
OY	384	GATTCGGGCACAGCTCAATTGCTTAGAGACCTCCAGACCGCGAAGCGCACTACC	443
Db	253	AATATAAAAAATCACTTCTATTTTGACTGAACCAATTCGCGGGCTTAAAAACGACTACCT	312
OY	444	CTGCTTGGCCCGATGAGCGAAGTGGCCGCTGACCTCGCCGCCAGTTGG-----	495
Db	313	ATTTTAGGGCTATGAGGGTGTGGTGGAGGTTGGCTGGCATTTAGTCAAGCATTAT	373
OY	495	-----GCTTACCACTGTATGCGAACCCAAAGGGGCGCGGTGCTGTGATGGCGGGGTG	548
Db	373	TTACTGGCTTTAGAGCATGTTAAAGGGTTTATGGGTAAAGGGGCTCATGCTAGAGGGGTTTG	432
OY	549	CCGGCGCTGCAACCGCGCAGTGTGTGTATCGGCGCGGCACCGCGCGCTTCAACAGCA	608
Db	433	TCGGGTGGCGAAAGGGCTAAATCTCTTAATTGAGGGCGGTGTGTTGGCATGGAGAGC	492
OY	609	GCCGCGATCGCCAAACGGCATGGCGCGCCACCGTTACGGTTCTAGACATCAACATGCACAA	668
Db	493	GCGAAAGCTTAACCAATGGGGGCTAAAGTAACGATTTAGAAATAGACTACGCTAAA	552
OY	669	CTTGGGCACTCGAAGCGCGAATTGTGGGCGCGCATCCACATCGCTACTCATGGGCTAC	728
Db	553	TTACAAAACCAACCTTATTAATCAATTTGTATGATTTAAGATCTTAAACGTGAATGAAGCC	612
OY	729	GAGCTCGAAGGTTGCCGCTCAACGTCGCCGACCTGTGTGATTTGGGCGCTCTGTGTCCAGGC	788
Db	613	AATATCATTCAGGCTTAAGAGGGCGCGTGGCGCTAGTGGAGCGGTGTGGTTACAGCG	672
OY	789	GCCAAAGGCAACCAATTATAGTGTCAATTCACTTGTGCGGCAATGAAACCAAGGTGGCGTA	848
Db	673	AGCCAAACCCCTAAAGTATCTTAAAGAGGATTTAAAAATATCATGACAGACACAAAGGGTA	722
OY	849	CTGCTGATATAGCCATCGACCAAGGGCGGCTGTTTCAAGAGGCTCAGACCGACCACTAC	908
Db	733	GTCATTGATGTGGCTTCAGTTTGGGGGGGCGCATTAAGAGCCATACGCCAAGACGCAT	792
OY	909	GACCAACCGAGCTTGGCGGTGCCAGACACGCTGTTTACTGCGTGGGGAACATGCCGCC	968
Db	793	TCCTAACCGGTGTATGTGGAAGAAAGTTTGTTCATTATGGCGTGGCGCAACATGCCAGGG	852

```

QY 969 TCGGTCCGAGACGTCGACCTACGCGCTGACCAACGCGACGATGCTGACG 1028
   || || || || || || || || || || || || || || || || || || ||
Db 853 ATTGTGCTAAACGACGCTTACGCGCTTATAGCCATGACGAGTGTCTTGTAT 912
QY 1029 CTGCGGACCATGCTGCGGG 1050
   || || || || || || || || || || || || || || || || || || ||
Db 913 TATTAGACCATGCTTGAAG 934

RESULT 9
V43039/c
V43039 standard; DNA; 544 BP.
AC V43039;
DE 09-NOV-1998 (first entry)
KW Streptococcus pneumoniae polypeptide coding region;
KM Polypeptide; ORF; open reading frame; infection; bacterial;
OS Streptococcus pneumoniae; diagnosis; prophylaxis; ds.
FH Key Location/Qualifiers
FT CDS complement (59..334)
FT /tag= a
FT /note= "polypeptide"
W0923631-AL.
PN 04-JUN-1998.
PF 24-NOV-1997; U21976.
PR 27-NOV-1996; US-031879.
PA (SMK ) SMITHKLINE BEECHAM CORP.
PI (SMK ) SMITHKLINE BEECHAM PLC.
PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO,
PI Reid RH, Zarlos PN;
PI WPI: 98-322654/28.
PI P-PSDB: W62769.
DR Streptococcus pneumoniae polynucleotides - useful for developing
PT products for diagnosis, prevention and treatment of infections e.g.
PT pneumonia, bacteraemia, meningitis or endocarditis
PS Claim 1, Page 165-166; 181p; English.
CC The sequence is that of a Streptococcus polypeptide coding region.
CC The polypeptide can potentially be used for the diagnosis and
CC prevention of bacterial infections, especially SP infection.
CC It may be used for the treatment of diseases such as otitis media,
CC conjunctivitis, pneumonia, bacteraemia, meningitis, sinusitis, pleural
CC empyema, endocarditis or infection of the cerebrospinal fluid.
SQ Sequence 544 BP; 149 A; 121 C; 124 G; 150 T;

Query Match 5.6%; Score 69.2; DB 1: Length 544;
Best Local Similarity 53.4%; Pred No. 2.6e-06;
Matches 191; Conservative 0; Mismatches 163; Indels 4; Gaps 2;

QY 690 TTCTGGGCGGATCCACATCTGCTACTATCGCGCTACGAGCTCGAGGTGCGTCAAA 749
   || || || || || || || || || || || || || || || || || || ||
Db 543 TTGGAAGTCAAAATTCATTTATATCATTAATCAATCAATGACAGATATGAGA 484
QY 750 CGTCCGACCTGTGATGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 809
   || || || || || || || || || || || || || || || || || || ||
Db 483 GATCGATGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 424
QY 810 TCAATTCATCTGCGGATGATGAACGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 869
   || || || || || || || || || || || || || || || || || || ||
Db 423 ACGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 365
QY 870 CAGGCGGCGCTTTTGA--AGGCTACGACGACGACGACGACGACGACGACGACGACG 926
   || || || || || || || || || || || || || || || || || || ||
Db 364 CAGGCGGCGCTTTTGAACGATGACGACGACGACGACGACGACGACGACGACGACG 305
QY 927 GTGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 966
   || || || || || || || || || || || || || || || || || || ||
Db 304 AAACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 245
QY 987 ACCGACGCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1044
   || || || || || || || || || || || || || || || || || || ||
Db 244 ACCATGCGCTTACCAATGCTACTCTTCTTATATGCAAGCTTTGGCTGCAAGAT 187

```

```

RESULT 10
V74370
ID V74370 standard; DNA; 31096 BP.
AC V74370;
DE 16-MAR-1999 (first entry)
KW Staphylococcus aureus confg SEQ ID #59.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc_feature 1201..1260
FT /tag= a
FT /note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FT misc_feature 3001..3060
FT /tag= b
FT /note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FT misc_feature 4801..4860
FT /tag= c
FT /note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FT misc_feature 6501..6560
FT /tag= d
FT /note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FT misc_feature 8401..8460
FT /tag= e
FT /note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FT misc_feature 10201..10260
FT /tag= f
FT /note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FT misc_feature 12001..12060
FT /tag= g
FT /note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FT misc_feature 13801..13860
FT /tag= h
FT /note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FT misc_feature 15601..15660
FT /tag= i
FT /note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FT misc_feature 17401..17460
FT /tag= j
FT /note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"

```







Fri Jun 23 09:31:36 2000

us-09-362-485-7.rng

Page 12

Search completed: June 22, 2000, 15:10:02  
Job time: 17928 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:11:10 ; Search time 341.15 Seconds  
(without alignments)  
468.2/4 Million cell updates/sec

Title: US-09-362-485-7

Perfect score: 1229

Sequence: 1 ATCTGCAGATTAATCGAAC.....GAGTAAGGAGGATGATG 1229

Scoring table: IDENTITY\_NMC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 6492525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/PTCUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/1na/Dackfil1seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45.4	3.7	893	US-08-403-852D-3	Sequence 3, Appl1
2	45.4	3.7	5392	US-08-403-852D-1	Sequence 1, Appl1
3	44.4	3.6	2582	US-08-816-105A-2	Sequence 2, Appl1
4	43.6	3.5	28958	US-08-258-261B-6	Sequence 6, Appl1
5	43.6	3.5	28958	US-08-456-837-6	Sequence 6, Appl1
6	43.6	3.5	28958	US-08-457-342-6	Sequence 6, Appl1
7	43.6	3.5	28958	US-08-457-666A-6	Sequence 6, Appl1
8	43.6	3.5	28958	US-08-458-076A-6	Sequence 6, Appl1
9	43.6	3.5	28958	US-08-764-233A-4	Sequence 4, Appl1
10	43.6	3.5	28958	US-08-457-335A-6	Sequence 6, Appl1
11	43.6	3.5	28958	US-08-729-214-6	Sequence 6, Appl1
12	43.6	3.5	49377	US-08-764-233A-1	Sequence 1, Appl1
13	42.4	3.4	1833	US-08-403-852D-6	Sequence 6, Appl1
14	42.4	3.4	2185	US-08-173-508-3	Sequence 3, Appl1
15	42.4	3.4	2185	US-08-265-310-3	Sequence 3, Appl1
16	41.6	3.4	1620	US-08-461-775-10	Sequence 10, Appl1
17	41.6	3.4	20235	US-07-642-734C-3	Sequence 3, Appl1
18	40.6	3.3	4257	US-08-690-473-1	Sequence 1, Appl1
19	40.6	3.3	12001	US-08-458-568A-11	Sequence 11, Appl1
20	39.6	3.2	2414	US-08-461-775-11	Sequence 11, Appl1
21	39.6	3.2	2668	US-08-461-775-11	Sequence 11, Appl1
22	39.4	3.2	459	US-08-387-942C-35	Sequence 35, Appl1
23	39.4	3.2	1998	US-08-387-942C-35	Sequence 35, Appl1
24	39.4	3.2	12588	US-08-387-942C-1	Sequence 1, Appl1
25	39.2	3.2	4420	US-08-470-119-148	Sequence 148, App
26	39.2	3.2	44377	US-08-804-227C-7	Sequence 7, Appl1
27	39.2	3.2	44377	US-08-804-198-1	Sequence 1, Appl1

28	39	3.2	43280	US-08-804-227C-1	Sequence 1, Appl1
29	38.8	3.2	8051	US-08-576-626A-2	Sequence 2, Appl1
30	38.8	3.2	11219	US-07-642-734C-1	Sequence 1, Appl1
31	38.6	3.1	2064	US-08-343-428-1	Sequence 1, Appl1
32	38.6	3.1	11219	US-07-642-734C-1	Sequence 1, Appl1
33	38.4	3.1	1215	US-08-947-726A-1	Sequence 1, Appl1
34	38.4	3.1	2048	US-08-776-251-1	Sequence 1, Appl1
35	38.2	3.1	2109	US-08-555-568B-20	Sequence 20, Appl1
36	38	3.1	474	US-08-403-852D-14	Sequence 14, Appl1
37	38	3.1	30001	US-08-125-468-1	Sequence 1, Appl1
38	38	3.1	30001	US-08-474-933-1	Sequence 1, Appl1
39	37.8	3.1	3252	US-08-809-740A-1	Sequence 1, Appl1
40	37.8	3.1	3252	US-08-809-740A-4	Sequence 4, Appl1
41	37.8	3.1	15664	US-08-402-282-3	Sequence 3, Appl1
42	37.8	3.1	15664	US-08-508-004-3	Sequence 3, Appl1
43	37.8	3.1	15664	US-08-402-066-3	Sequence 3, Appl1
44	37.8	3.1	15664	US-08-402-068-3	Sequence 3, Appl1
45	37.6	3.1	1524	US-08-402-068-3	Sequence 3, Appl1
				5512669-1	Patent No. 5512669

#### ALIGNMENTS

RESULT 1  
US-08-403-852D-3  
Sequence 3, Application US/08403852D  
Patent No. 5891695  
GENERAL INFORMATION:  
APPLICANT: Blanc, Veronique  
APPLICANT: Blanchet, Francis  
APPLICANT: Crouzet, Joel  
APPLICANT: Jacques, Nathalie  
APPLICANT: Lacroix, Patricia  
APPLICANT: Thibaut, Denis  
APPLICANT: Zagorec, Monique  
APPLICANT: Debussche, Laurent  
TITLE OF INVENTION: De Grey-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved in the  
TITLE OF INVENTION: Synthesis of Streptogramins, Nucleotide Sequences  
TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Finnegan, Henderson, Farbow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,852D  
FILING DATE: 10-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806.0054-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4400  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 833 base pairs  
TYPE: nucleic acid





COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage  
 COMPUTER: Apple Macintosh 6500  
 OPERATING SYSTEM: System 7.5  
 SOFTWARE: WordPerfect 3.5  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/816,105A  
 FILING DATE: 14-MAR-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: DE 19625269.5  
 FILING DATE: 25-JUN-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: DE 19611252.4  
 FILING DATE: 22-MAR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kurt G. Brigg  
 REGISTRATION NUMBER: 33,141  
 REFERENCE/DOCKET NUMBER: Bayer 9814-KGB  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (914) 332-1700  
 TELEFAX: (914) 332-1844  
 INFORMATION FOR SEQ. ID NO. 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2587 nucleotides  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 US-08-816-105A-2

Query Match 3.6%; Score 44.4; DB 4; Length 2582;  
 Best Local Similarity 46.9%; Pred. No. 0.031;

Matches 138; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 352 CCCGCTACCTGCTTGCACGATGCGTGTGATTCGCGACACGATCAATTCCTACG 411  
 DB 1576 CCCGCTACCTGCTTGCACGATGCGTGTGATTCGCGACACGATCAATTCCTACG 1635  
 QY 412 AGACCGTACGACGCGGCGGACGACGACGACGACGACGACGACGACGACGACGACG 471  
 DB 1636 TGCACGCTGACGACGCGGCGGACGACGACGACGACGACGACGACGACGACGACGACG 1695  
 QY 472 GTGACGCTGCGCGCGGCGGACGACGACGACGACGACGACGACGACGACGACGACG 531  
 DB 1696 GCGACGCTGCGCGCGGCGGACGACGACGACGACGACGACGACGACGACGACGACG 1755  
 QY 532 TCGTGTGACGCGCGGCGGACGACGACGACGACGACGACGACGACGACGACGACG 591  
 DB 1756 TCGGCGCGGCTGTCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1815  
 QY 592 CCGCGCGCTACACGACGCGGCGGACGACGACGACGACGACGACGACGACGACGACG 645  
 DB 1816 CCGCGCGACGCTGCGCGATGCGCGAGTCCCGCGACCGTCCAGCATGAGACG 1869

RESULT 4

US-08-258-261B-6

Sequence 6, Application US/08258261B

Patent No. 5639949

GENERAL INFORMATION:

APPLICANT: Schnupp, Thomas

APPLICANT: Lilgon, James M.

APPLICANT: Beck, James Joseph

APPLICANT: Hill, Dwight Steven

APPLICANT: Ryals, John Andrew

APPLICANT: Gafney, Thomas Deane

APPLICANT: Lam, Stephen Ting

APPLICANT: Hammer, Phillip E.

APPLICANT: Uknes, Scott Joseph

TITLE OF INVENTION: Genes for the synthesis of

TITLE OF INVENTION: antipathogenic substances

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba-Geigy Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/258,261B

FILING DATE: 08-JUN-1994

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/457,205

FILING DATE: 01-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129

REFERENCE/DOCKET NUMBER: CGC 1506/CIP3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8614

TELEFAX: 919-541-8689

INFORMATION FOR SEQ. ID NO. 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 28958 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-258-261B-6

Query Match 3.5%; Score 43.6; DB 1; Length 28958;

Best Local Similarity 44.8%; Pred. No. 0.096;

Matches 290; Conservative 0; Mismatches 354; Indels 4; Gaps 3;

QY 148 GGCATGAGTGTCTTCATCCAGGAGGCGGAGAGGCGTGGCTATCACGACGGGAT 207  
 DB 25306 GGCATGAGTGTCTTCATCCAGGAGGCGGAGAGGCGTGGCTATCACGACGGGAT 25365  
 QY 208 TTCAAGGCGGAGCGCGGCGGACGACGACGACGACGACGACGACGACGACGACGACG 267  
 DB 25366 TTCAAGGCGGAGCGCGGCGGACGACGACGACGACGACGACGACGACGACGACGACG 25425  
 QY 268 TTATGCTCAAGGTCAA-AGAACGATAGCGGCGGATACGCGCGCTCGACAGGGCG 326  
 DB 25426 TTATGCTCAAGGTCAA-AGAACGATAGCGGCGGATACGCGCGCTCGACAGGGCG 25485  
 QY 327 ATCTGTACGCTTCTTGTGATTTGGCGCGGCTGACGCTTGCACCGATCGTTGGAT 386  
 DB 25486 ATCTGTACGCTTCTTGTGATTTGGCGCGGCTGACGCTTGCACCGATCGTTGGAT 25545  
 QY 387 TCGGACACGATGATTCGCTACGACGCGGCGGACGCGGCGGCGGCGGCGGCGGCT 446  
 DB 25546 TCGGACACGATGATTCGCTACGACGCGGCGGACGCGGCGGCGGCGGCGGCGGCT 25605  
 QY 447 CTGCGCGGATGAGGAGGCGGCGGCTGACGCTGCGCGCGGCGGCGGCGGCGGCTTAC 504  
 DB 25606 CTGCGCGGATGAGGAGGCGGCGGCTGACGCTGCGCGCGGCGGCGGCGGCGGCTTAC 25665  
 QY 505 TGATGGAACCAAGGCGGCGGCGGCTGCTGATGCGGCGGCGGCGGCGGCGGCTGAC 564  
 DB 25666 TGATGGAACCAAGGCGGCGGCGGCTGCTGATGCGGCGGCGGCGGCGGCGGCTGAC 25725  
 QY 565 CCGACG-TGCTGTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 623  
 DB 25726 CCGACG-TGCTGTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 25785



```

1  COMPUTER: IBM PC compatible
2  OPERATING SYSTEM: PC-DOS/MS-DOS
3  SOFTWARE: PatentL Release #1.0, Version #1.25
4
5  CURRENT APPLICATION NUMBER: US/08/457,342
6  APPLICATION NUMBER: US/08/457,342
7  FILING DATE: 01-JUN-1995
8  CLASSIFICATION: 424
9
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER: US 08/457,205
12 FILING DATE: 01-JUN-1995
13 APPLICATION NUMBER: 08/258,261
14 FILING DATE: 08-JUN-1994
15 ATTORNEY/AGENT INFORMATION:
16 NAME: Elmer, James Scott
17 REGISTRATION NUMBER: 36,129
18 REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
19 TELECOMMUNICATION INFORMATION:
20 TELEPHONE: 919-541-8614
21 TELEFAX: 919-541-8689
22 INFORMATION FOR SEQ ID NO: 6:
23 SEQUENCE CHARACTERISTICS:
24 LENGTH: 28958 base pairs
25 TYPE: nucleic acid
26 STRANDEDNESS: single
27 TOPOLOGY: linear
28 MOLECULE TYPE: DNA (genomic)
29 HYPOTHEtical: NO
30 ANTI-SENSE: NO
31
32 US-08-457-342.6

```

Db	25366	TTCCGAGCACCGGCGCGCTCTCGTACCCACAAACCGGAGACGACTCTCTCCGCGCTCGAC	25425
QY	268	TTATATGCTCAAGTCAA-AGAACCGATAGCGCGGAAATACGGCCCTTCGCACACGGGCG	326
Db	25426	TCCGTCGCCGCCAAGACAAAGCCCCCGGACACCGTCTCGGACGAGCGAAGCCACGGC	25485
QY	327	ATCTGTGTCACGTTCTTGCAATTTGGCGCGGCTACAGCGCTTCGACCGATGCGTTGTTGAT	386
Db	25486	AAGCTGCTTCTTCCTTCCTTCCTGGGACAGGCTTCGACGTGGGAAGGATGAGCCCTTCGCTG	25545
QY	387	TCCGCGACACGACTCAATTGCTTACGAGACCGTCCAGACCGCCGAAAGCGCAGCTACCCCTG	446
Db	25546	CTCGACTCTCCGCTCCGCTCTTTCGGCGCTCAGCTCGAAGCATGGGACGCGCGCTCCTCT	25605
QY	447	CTTGCCCCGATAGCGAATCGCCGCTGCACATCGCCGCGAGTTGGGCTTACAC- -C	504
Db	25606	CACGTCGATGGAGCTGCTCGCCCGCTTCGCGCCCGGAGAGAGGGGCCCTTCCTCCGAC	25665
QY	505	TATATGCAACCCAAAGGGGCGCGGATGTCCTATGGGCGGGGTGCCCGGCTCGAACCG	564
Db	25666	CCCGTCGACGCTGTACAGCCGCCCGCTTTGGCGATGATGTCCTCCCTGGCGGCGCTCTGG	25725
QY	565	CCGAGC-TCGTGTGATCGGCGCGCGGACACCGCGGCTACAAAGCAGACCCGATCGCAAC	623
Db	25726	CCCTCGCTCTGGGGTAAAGCCCGCCCGCGTGTCTGGGCAACAGTCAGGGGCGAGATCGCGCC	25785
QY	624	GGCATGGGGCGCAGCGCTTACGAGTTCTAGACATCAACATGACAAATTGGCAACTCGAC	683
Db	25786	GCTTTCGTGCGAGGCGCTCTCTCCCTCGAAGAGGCGGCGCCCGATCGCCCGCTGGGCAC	25845
QY	684	GCCGAGTTCGCGGCGCGATCCACACACTGCTACTATCGGCTACGAGCTCGAGGTGCC	743
Db	25846	AAAGCGCTACACACGCTCCCGGCAACGGGGCATATGGC- -JCGCTCAGCTCGCGCGCTCC	25905
QY	744	GTCMAACGTCCGCACTGTGATTTGGGGGCGTCTGTCGCAAGGCGC- -791	
Db	25906	GACCTTCAGACTACTCTGCTGCTCTTGGGCGACAGGCTCTTCATCGCC- -25953	



REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-541-8614  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 28958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 US-08-457-335A-6

Query Match 3.5%; Score 43.6; DB 2; Length 28958;  
 Best Local Similarity 44.8%; Pred. No. 0.096;  
 Matches 290; Conservative 0; Mismatches 354; Indels 4; Gaps 3;

148 GGCCTAGAGTGTCTATCCAGGAGGCTCCGAGAGGCTCCGCTTACACGAGCGGAT 207  
 25306 GCCCAGACACCTCAGCCCTCGGGGATGTGGCTTATTCCTGACACCCGCGCCAC 25365  
 208 TTCAAGGCGGACGAGCGGCACTGTGCGACCGCGGAGGCTGTGGGCGACGCTGAT 267  
 25366 TTGAGAGACCGGCGCCCTCTCGTAGCCACACACCGGAGAGGCTCTCTCCGCGTCGAC 25425  
 268 TTATTGCTCAAGTCAA-AGAACCGATAGCGGGGAAATACGGCCGCTCGACACGGGG 326  
 25426 TCAGTCGCCAGAGCAAGCCCGCCGAGACCTCTCTGAGGAGGAGCA] CACGGC 25485  
 327 ATCTTGTCAGTTCAGGATTTGGCGGCGGCTGACGTGTTGACCGATGGTGTGGAT 386  
 25486 AACCTGCT 25545  
 387 TCCGACACACGTCATATGCTTACGAGACCGTCCAGACCGCGGAGCGGACATACCCCTG 446  
 25546 CTCGACTCTCCGCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25605  
 447 CTGCGCCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 504  
 25606 CAGCTGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 25665  
 505 TGATGGGAAACCAAGGGGGGCGGAGTGTGCTGATGGGCGGGGCGGCGGAGTGAACCG 564  
 25666 CCGCTGAGAGCTGACAGCGCGCGCTCTTGGCGTATGTGTCTCTCTCTCTCTCTCT 25725  
 565 CCGAGC-TGCTGTGATCGGGCGGCGGCGGCGGCTACACAGGAGCGGCGATCGGCAAC 623  
 25726 CGCTCGCTGGCTAGAGCGCGCGGCTGTGTCGACAGTGAAGGCGAGATCGCGCGC 25785  
 624 GGCATGGGCGGAGCGGCTTACGCTTACGATACACATGACAAATTGGCAACTCGAC 683  
 25786 GCCTTGTGCGAGGCGCT 25845  
 684 GCGGATTTGCGGCGGAGTCCACACTCGCTACTATCGGCTTACGAGCTCGAGGGTGGC 743  
 25846 AAGCGCTTACACACCTCCCGGCAAGGGGCGGCGGCGGCTCGAGCTCGGCGCTCC 25905  
 744 GTCAAGCTGCGGAGCTGTGATTTGGGCGGCTGCTGTGCGGAGGCGC 791  
 25906 GACCTCAGACCTACT 25953

RESULT 11  
 US-08-729-214-6  
 Sequence 6, Application US/08729214  
 Patent No. 5817502  
 GENERAL INFORMATION:  
 APPLICANT: Ligon, James M.  
 APPLICANT: Hall, Dwight Steven  
 APPLICANT: Ryals, John Andrew

APPLICANT: Hammer, Phillip E.  
 APPLICANT: van Pee, Karl-Heinz  
 APPLICANT: Kirner, Sabine  
 TITLE OF INVENTION: Genes for the synthesis of  
 TITLE OF INVENTION: antipathogenic substances  
 NUMBER OF SEQUENCES: 27  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Clida-Gelgy Corporation  
 STREET: 520 White Plains Road  
 CITY: Tarrytown  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10591  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/729, 214  
 FILING DATE: TBA  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meigs, J. Timothy  
 REGISTRATION NUMBER: 38,241  
 REFERENCE/DOCKET NUMBER: CGC 1506/CIP5  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-541-8587  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 28958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 US-08-729-214-6

Query Match 3.5%; Score 43.6; DB 2; Length 28958;  
 Best Local Similarity 44.8%; Pred. No. 0.096;  
 Matches 290; Conservative 0; Mismatches 354; Indels 4; Gaps 3;

148 GGCCTAGAGTGTCTATCCAGGAGGCTCCGAGAGGCTCCGCTTACACGAGCGGAT 207  
 25306 GCCCAGACACCTCAGCCCTCGGGGATGTGGCTTATTCCTGACACCCGCGCCAC 25365  
 208 TTCAAGGCGGACGAGCGGCACTGTGCGACCGCGGAGGCTGTGGGCGACGCTGAT 267  
 25366 TTGAGAGACCGGCGCCCTCTCGTAGCCACACACCGGAGAGGCTCTCTCCGCGTCGAC 25425  
 268 TTATTGCTCAAGTCAA-AGAACCGATAGCGGGGAAATACGGCCGCTCGACACGGGG 326  
 25426 TCAGTCGCCAGAGCAAGCCCGCCGAGACCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25485  
 327 ATCTTGTCAGTTCAGGATTTGGCGGCGGCTGACGTGTTGACCGATGGTGTGGAT 386  
 25486 AACCTGCT 25545  
 387 TCCGACACACGTCATATGCTTACGAGACCGTCCAGACCGCGGAGGAGGAGGAGGAGG 446  
 25546 CTCGACTCTCCGCGCCCT 25605  
 447 CTGCGCCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 504  
 25606 CAGCTGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 25665  
 505 TGATGGGAAACCAAGGGGGGCGGAGTGTGCTGATGGGCGGGGCTCCGCGGCTGGAACCG 564  
 25666 CCGCTGAGAGCTGACAGCGCGCGCTCTTGGCGTATGTGTCTCTCTCTCTCTCTCTCT 25725



QY	565	CCGAGC-TGCTGTATCGGCGCGGACCGCGGCTTACACGACCGCGATCGGCAC	623
Db	25726	CGCTGCTCGGGTAGAGCCGCCGCCCGCTGTCGCGCACATGACGCGCATCGCGCC	25785
QY	624	GGCATGGCGCGACCGATTACGGTTCTAGACATCAACATGACACAACTTCGGCAATCGAC	683
Db	25786	GCGTTCTGTCGAGGGCGTCTCTCCCTCGAGAGCGGGCCCGCATGCGCCCGCTGCGCAGC	25845
QY	684	GCCGAGTCTGGGGCGGATTCACACTCGCTACTCATCGCGCTAGAGTTCAGAGGTGCC	743
Db	25846	AAAGGCGTACACCTGTCGCGCGGACACGGGCGCATGCGCGCGTCTGAGTCTGCGCGCTCC	25905
QY	744	GTCAAACGTCGCGACCTGTGATTGGGCGCGCTGCTGCTGTCACAGCGCC	791
Db	25906	GACCTCGACACTTACTCGCTCCCTCGGGGCGACAGGCTCTCCATCGCC	25953
RESULT 12			
US-08-764-233A-1			
Sequence 1, Application US/08764233A			
Patent No. 5716849			
GENERAL INFORMATION:			
APPLICANT: Ligon, James M.			
APPLICANT: Schnupp, Thomas			
APPLICANT: Beck, James J.			
APPLICANT: Hill, Dwight S.			
APPLICANT: Neff, Suzanna			
APPLICANT: Ryals, John A.			
TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen			
NUMBER OF SEQUENCES: 10			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Ciba-Geigy Corporation			
STREET: 520 White Plains Road, P.O. Box 2005			
CITY: Tarrytown			
STATE: NY			
COUNTRY: USA			
ZIP: 10591			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: Patent In Release #1.0, Version #1.30			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/764,233A			
FILING DATE:			
CLASSIFICATION: 435			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 08/729,214			
FILING DATE: 09-OCT-1996			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 08/258,261			
FILING DATE: 08-JUN-1994			
ATTORNEY/AGENT INFORMATION:			
NAME: Meigs, J. Timothy			
REGISTRATION NUMBER: 38,241			
REFERENCE/DOCKET NUMBER: 1506/CIP6			
TELEPHONE: (919) 541-8587			
TELEFAX: (919) 541-8689			
INFORMATION FOR SEQ ID NO: 1:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 49377 base pairs			
TYPE: nucleic acid			
STRANDEDNESS: single			
TOPOLOGY: linear			
MOLECULE TYPE: DNA (genomic)			
ORIGINAL SOURCE:			
ORGANISM: Sorangium cellulosum			
IMMEDIATE SOURCE:			
CLONE: p98/1, pU13, and pYKM15			
FEATURE:			
NAME/KEY: misc.feature			
LOCATION: 383..760			

```

OTHER INFORMATION: /product= "SorA"
/note= "This gene encodes a protein that is highly homologous to type I PRSs such as eryA from Saccharopolyspora erythraea."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 927..19874
OTHER INFORMATION: /product= "SorA"
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PRSs t
OTHER INFORMATION: are known to be involved in the synthesis of polypeptide
OTHER INFORMATION: compounds."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 942..7115
OTHER INFORMATION: /product= "Module 1 of SorA"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7203..12884
OTHER INFORMATION: /product= "Module 2 of SorA"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 13455..19616
OTHER INFORMATION: /product= "Module 3 of SorA"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19871..46318
OTHER INFORMATION: /product= "SorB"
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PRS ge
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19870..24556
OTHER INFORMATION: /product= "Module 1 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 24638..30820
OTHER INFORMATION: /product= "Module 2 of SorB"
FEATURE:
NAME/KEY: misc_featu
LOCATION: 30881..354
OTHER INFORMTION: /product= "Module 3 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 35528..40114
OTHER INFORMATION: /product= "Module 4 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 40190..46318
OTHER INFORMATION: /product= "Module 5 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 46851..47891
OTHER INFORMATION: /product= "SorM"
OTHER INFORMATION: /note= "The protein encoded by the sorM gene is highly
OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the
OTHER INFORMATION: polyketide rapamycin."
US-08-764-233A-1

Query Match      3.58; Score 43.6; DB 2; Length 49377;
Best Loc: Similarity 44.88; Pred. No. 0.11;
Matches 290; Conservative 0; Mismatches 354; Indels 4; Gaps 3;

QY 148 GCCCATGAGTGGTCATCCACGACGAGTGCGGAGAGGGCTCGGTATCACCAGCGCGGAT 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41609 GCCCACGACGACTACCCCTCGCGGATGTGCCATTTCGCTGGCGACACACCGCGCCAC 41668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 208 TTCAAGGCGGACGCGCACTGTCGGACCGCGCACCAAGTGTGGCCGACGCTGAT 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41669 TTCGAGCACCGCGCGCTCTGTGATGCCACACCGCGACGAGCTCTCTCGCGCTGCAC 41728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

ZY 268 TTATGGCTCAAGTGTCAA-AGAACGATAGCGGCGGGAATAGGCGCGCTGGGACACGGCGG 326

```

```

Db 41729 TCGCTGCGCCAGACAAAGCCGCCGAGACCGCTCTCGAGAGGAGCGGAAGCCAGGC 41788
Qy 327 ACTTTGTCAGCTTCTTGATTTGGCCGGCTGACGTGTCAGCCAGTGGTTGTGAT 386
Db 41789 AAGCTGTCTTGTCTTCTTGCGGCAAGGCTGCAATGGGAAGGATGGCCCTCTGCTG 41848
Qy 387 TCCGGACCAAGCTCAATGCTTCTAGAGACCGTCCAGACCGCGGAAGCGCACTACCCCTG 446
Db 41849 CTCGACTCTCTGCGCCCTCTGCGCGCTGAGCTGGAAGCATGGAGCGCGCGCTGCTCT 41908
Qy 447 CTGCGCCGATGAGCGAGTTCGCGCGCTGCACTGCGCCGAGTTGGGCTTACCAAC--C 504
Db 41909 CAGCTGATGAGAGCTGCTGCTGCGCTGCTGCGCGAGAGAGGCGCCCTCTCTGAG 41968
Qy 505 TGTATCGCAACCAAGAGGCGCGCGCTGCTGATGGCGGGGTGCGCGCGTGAACCG 564
Db 41969 CCGCTGAGCTGATGACGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 42028
Qy 565 CCGAGC-TGCTGATGATGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 623
Db 42029 CCGCTGCTGCGCGCTGAGAGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 42088
Qy 624 GGCATGCGCGCGCGCGCTGATGCTTCTAGACATCAACATCAACAACTTGGCACTGAGC 683
Db 42089 GCGCTGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 42148
Qy 684 GCGGAGTTCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 743
Db 42149 AAGAGCTTACCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 42208
Qy 744 GTCAAACTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 791
Db 42209 GACCTTACCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 42256

RESULT 13
US-08-403-852D-6
Sequence 6, Application US/08403852D
Patent No. 5891695
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanche, Francis
APPLICANT: Crozet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Crecy-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852D
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441

```

```

FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4400
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 103..1689
US-08-403-852D-6

Query Match
Best Local Similarity 3.48; Score 42.4; DB 3; Length 1833;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Qy 532 TGCATGAGGGGGGGTGGCCGCGCTGCAACCGCGCGAGCTGTGTATGCGCGCGCA 591
Db 413 TGAATCTGGAGTGTGCTTCTGCGCGCGCGCGCGCTGCGCGCGCTGCGCGAGACT 472
Qy 592 CCGCGCGCTACACAGAGCGCGCGATGCGCAAGCGCATGGCGCGCGCGCTTACG 651
Db 473 CCGCGCACTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCA 532
Qy 652 ACATCAATGACAAACTTGGCACTGACGCGCGAGTTGCGCGCGATCCACACTC 711
Db 533 CCACCTTGCCTGATCATGAGGCTCTGCGAGCAGCAAGCGCGCGCGCGCGCGCG 592
Qy 712 GCTACTGCGCGCTACGA,CTCG 735
Db 593 TGTGGGAGCGCGCGCGCGCTGCG 516

RESULT 14
US-08-173-508-3
Sequence 3, Application US/08173508
Patent No. 5616485
GENERAL INFORMATION:
APPLICANT: Bartfield, Daniel
APPLICANT: Butler, Michael J.
APPLICANT: Hadary, Dany
APPLICANT: Jenish, David
APPLICANT: Krieger, Timothy
TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/173,508

```

```

FILING DATE: 23-DEC-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 18740/125/CACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2185 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 531..2069
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 531..902
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 903..2069
FEATURE:
NAME/KEY: misc_feature
LOCATION: 531..533
OTHER INFORMATION: /note="Met at position -124"
OTHER INFORMATION: represents fMet"
US-08-173-508-3

```

```

Query Match          3.4%; Score 42.4; DB 1; Length 2185;
Best Local Similarity 47.4%; Pred. No. 0.091;
Matches 127; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```

```

QY 524 CCGCGGTGTGCTGATGAGGCGGGGTCCCGGGGTGACACCGCCGACGCGGTGATCGG 583
DB 1250 CCGCGTCCGCGATGACCGCGGCGTGACACCGCGGGTGAAGTGGCCGATCAAGTCTCCAA 1309
QY 584 CCGCGGACCGCGCGGTCTACCAACGACCGCGATCGCCCAACGCGATGGCGGACCGTTAC 643
DB 1310 CCGCGACGGCTCTTCTACACCGAGCGGTGATGCGGCTTCACTTCACTCACTGCAAGA 1369
QY 644 GGTTCAGACATCAACATCGCAAACTTCGGCACTCGACCGCGAGTCTCGCGCCGAT 703
DB 1370 CGGCGTCGACGTGACCAACACAGCTATTACACCGACCGCGTGTCACTCACTGCAAGA 1429
QY 704 CCACACTCGCTACTCATGAGCGCTACGAGCTCGAGGCTCGCTCAAACTGCGGACCTGT 763
DB 1430 CGACCCGACACGAGGCGGTCTGAGAGCGGTCTCGCGGGCTTCGCGGTACGCGGAGAA 1489
QY 764 GATTGGGCGCTCTGTTGTCAGCGGCC 791
DB 1490 GAAGGCGCGGTCAACGTCGCCGCGCC 1517

```

```

RESULT 15
US-08-265-310-3
Sequence 3, Application US/08265310
Patent No. 5856166
GENERAL INFORMATION:
APPLICANT: Bartfeld, Daniel
APPLICANT: Butler, Michael J.
APPLICANT: Hadary, Dany
APPLICANT: Jenish, David
APPLICANT: Krieger, Timothy
APPLICANT: Malek, Lawrence T.
APPLICANT: Soostmeyer, Gisela
APPLICANT: Walczyk, Ewa
APPLICANT: Kryszman, Phyllis
APPLICANT: Garven, Sheila

```

```

TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESSES:
ADDRESS: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/265,310
FILING DATE: 24-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,508
FILING DATE: 23-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 18740/133/CACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2185 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 531..2069
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 531..902
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 903..2069
FEATURE:
NAME/KEY: misc_feature
LOCATION: 531..533
OTHER INFORMATION: /note="Met at position -124"
OTHER INFORMATION: represents fMet"
US-08-265-310-3

```

```

Query Match          3.4%; Score 42.4; DB 3; Length 2185;
Best Local Similarity 47.4%; Pred. No. 0.091;
Matches 127; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```

```

QY 524 CCGCGGTGTGCTGATGAGGCGGGGTCCCGGGGTGACACCGCCGACGCGGTGATCGG 583
DB 1250 CCGCGTCCGCGATGACCGCGGCGTGACACCGCGGGTGAAGTGGCCGATCAAGTCTCCAA 1309
QY 584 CCGCGGACCGCGCGGTCTACCAACGACCGCGATCGCCCAACGCGATGGCGGACCGTTAC 643
DB 1310 CCGCGACGGCTCTTCTACACCGAGCGGTGATGCGGCTTCACTTCACTCACTGCAAGA 1369
QY 644 GGTTCAGACATCAACATCGCAAACTTCGGCACTCGACCGCGAGTCTCGCGCCGAT 703
DB 1370 CGGCGTCGACGTGACCAACACAGCTATTACACCGACCGCGTGTCACTCACTGCAAGA 1429
QY 704 CCACACTCGCTACTCATGAGCGCTACGAGCTCGAGGCTCGCTCAAACTGCGGACCTGT 763
DB 1430 CGACCCGACACGAGGCGGTCTGAGAGCGGTCTCGCGGGCTTCGCGGTACGCGGAGAA 1489

```

Fri Jun 23 09:31:37 2000

us-09-362-485-7.rni

Page 12

QY      764 GATTGGGCGCTGCTGTGCCAGCGCC 791  
         || | | | | | | | |  
Db    1490 GAAGGGCGCGTCAACGTGCGCGCGCC 1517

Search completed: June 22, 2000, 15:13:36  
Job time: 18294 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:07:43 ; Search time 5541.94 seconds  
(without alignments)  
898.857 Million cell updates/sec

Title: US-09-362-485-7  
Perfect score: 1229  
Sequence: 1 ATCTTGACATTAATCAAC.....GAGTAAGGAGACGATGATG 1229

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 202611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_est48:\*  
83: gb\_est49:\*  
84: gb\_est50:\*  
85: gb\_est51:\*  
86: em\_est34:\*  
87: em\_est35:\*  
88: em\_est36:\*  
89: em\_est37:\*  
90: gb\_est52:\*  
91: gb\_est53:\*  
92: gb\_est54:\*  
93: gb\_est55:\*  
94: gb\_est56:\*  
95: em\_est38:\*  
96: em\_est39:\*  
97: em\_est40:\*  
98: em\_est41:\*  
99: em\_est42:\*  
100: em\_est43:\*  
101: em\_est44:\*  
102: gb\_est57:\*  
103: gb\_est58:\*  
104: em\_est45:\*  
105: gb\_est59:\*  
106: gb\_est60:\*  
107: gb\_est61:\*  
108: gb\_est62:\*  
109: gb\_est63:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES

No.		Score	Match length	DB	ID	Description
C	1	57.4	4.7	925	82	CNS0091P
	2	54	4.4	925	82	CNS0091P
	3	51.8	4.2	535	74	AV390505
C	4	49	4.0	910	82	CNS0060N
	5	46.6	3.8	846	82	CNS0108J
	6	45.4	3.7	591	69	AM128786
C	7	45.4	3.7	844	82	CNS0055P
	8	45.4	3.7	935	82	CNS0063K
	9	45.2	3.7	1101	83	CNS0178Y
C	10	45	3.7	315	51	AT171257
	11	45	3.7	846	82	CNS0108J
	12	45	3.7	1101	83	CNS0181E
C	13	44.8	3.6	932	82	CNS0072Q
	14	44.2	3.6	935	82	CNS0067K
	15	43.8	3.6	427	51	AT176678
C	16	43.8	3.6	512	74	AV367794
	17	43.8	3.6	521	74	AV397116
	18	43.8	3.6	526	74	AV394827
C	19	43.8	3.6	744	83	CNS0127K
	20	43.6	3.5	1101	80	CNS0160E
	21	43.4	3.5	414	80	AW286084
C	22	43.4	3.5	465	80	AW287007
	23	43.2	3.5	439	51	AQ08405A
	24	43.2	3.5	798	82	CNS000AA
C	25	43	3.5	413	74	AV396947
	26	42.6	3.5	645	82	CNS01213
	27	42.4	3.4	843	82	CNS00331
C	28	41.8	3.4	530	74	AV396353
	29	41.8	3.4	932	82	CNS0072Q
	30	41.6	3.4	506	74	AM180713
C	31	41.6	3.4	870	82	CNS0066Z
	32	41.4	3.4	1101	82	CNS001KJ
	33	41.2	3.4	639	60	AT181214
C	34	41.2	3.4	774	105	AQ3272121
	35	41.2	3.4	910	82	CNS0060N
	36	41	3.3	605	46	AT161529
C	37	41	3.3	1101	83	CNS0178Y
	38	40.8	3.3	384	64	AM054773
	39	40.8	3.3	385	21	D48746
C	40	40.8	3.3	426	61	C91740
	41	40.8	3.3	437	74	AV395790
	42	40.8	3.3	839	82	CNS004AB
C	43	40.8	3.3	964	82	CNS003MG
	44	40.6	3.3	506	74	AV396405
	45	40.4	3.3	574	91	AQ0848186
						AL053013 Drosophila
						AL053013 Drosophila
						AV390505 AV390505
						AL065629 Drosophila
						AL099337 Drosophila
						AM128786 fe17d05.y
						AL056652 Drosophila
						AL066051 Drosophila
						AL108460 Drosophila
						AT171257 605069D09
						AL099337 Drosophila
						AL108764 Drosophila
						AL066742 Drosophila
						AL066051 Drosophila
						AT176678 sb32B04.y
						AV396794 AV396794
						AV397116 AV397116
						AV394827 AV394827
						AL06598 Drosophila
						AL107216 Drosophila
						AM286084 LG1_261_B
						AM287007 LG1_264_C</

collaboration with the BerkeleyDrosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see [http://www.fruitfly.org/TheBDGP/drosophila\\_melanogaster\\_BAC\\_library](http://www.fruitfly.org/TheBDGP/drosophila_melanogaster_BAC_library) was prepared by Kazutoyo Oosawa and Aaron Mammoser. In Pieter de Jong's laboratory in the Department of Cancer Genetics at the "Cooper" Park Cancer Institute in Buffalo, NY. The library is named RCGI-98 and was constructed by partial *NotI* digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```

FEATURES
  source
    Local:09/Qualifiers
      1..925
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone_1b="RPI.48"
        /clone="BACR19D16"
        /note-end="TET3"
      61 c      61 g      172 t      511 others
120 a
ORIGIN

```

Query Match 4.7%; Score 57.4; DB 82; Length 925;  
Best Local Similarity 11.9%; Pred. No. 0.056;  
Matches 43; Conservative 176; Mismatches 142; Indels 0; Gaps 0;

[illegible][illegible]







/sex="mixed"  
 /tissue\_type="26 somite embryos, adult livers, shield  
 stage embryos"  
 /lab\_host="XLI-blue MRF"  
 /note="vector: pSPORI1; site\_1: NotI; site\_2: SalI; 1st  
 strand cDNA was primed with a Not I - oligo(dT)15 primer  
 5'-pGACGATGCTGCTGATGCTGCGGCGCCGCCCTTTTCTTTTCTT-3';  
 double-stranded cDNA was ligated to Sal I adaptors (BRL),  
 digested with Not I and cloned into the Not I and Sal I  
 sites of the pSPORI1 vector (BRL). Library was constructed  
 by Matthew Clark (Lehrbach Lab; ICRF, London and Max Planck  
 Institut fuer Molekulare Genetik, Berlin). cDNAs for EST  
 analysis were selected following oligonucleotide  
 hybridization fingerprinting of arrayed clones from  
 zebrafish late somitogenesis (26 ss), adult liver or  
 embryonic shield stage (3, 5, 6 h) libraries. Fingerprint  
 data were used to computationally cluster cDNAs, and a  
 single cDNA from each cluster was chosen for sequencing.  
 In some cases multiple members of the same cluster were  
 sequenced to assess clustering parameters or single clones  
 were sequenced additional times to assess quality  
 control."

BASE COUNT  
ORIGIN

Query Match	3.7%	Score 45.4	DB 69	Length 591
Best Local Similarity	50.2%	Pred. No. 11		
Matches 112; Conservative	0	Mismatches 111	Indel 0	Gaps 0

QY	2	TCCTCAGATTATATGAAGTCTCTTACACCTATACGATACACTATACGAGAGGGGATATCA	61
QY	185	TCCTTATAGGACACATATAGAGCCCTCAACCGCGCTACTAGTTCAGGCAATTCTCTTACAAACAGC	24
Db	62	TCGCGGTGGGTATTCGAGCCGAGACCAAAACAGCAATTCGCGGTGGGCATACCCGG	12
QY	245	TCACGCTGGGGCTTCTCTAAGGAAATCTTTACAAATAGAGCTGAGATGGGCATCTCTCTG	30
Db	122	CCGCGCTGCGGAACTTAACCCGTCGTGGCCATGAGTGTCTATCCAGGAGAGTGGCCGAG	18
QY	305	CCGAGATGGAACTCTCATCAACGAGCGGCTTAACTGTGTGGAGTCTCGGCGCGGAG	36
Db	182	AGGCTCGGCTATCAACCAAGCGGATTTCAAGCGCGGACGGCC	224
QY	365	ATCTGCGCAAGTCTCTGATGACATGTATACAAAAGCAGAGAC	407

RESULT	7				
LOCUS	CNS0052P/c				
DEFINITION	CNS0052P	844 bp	DNA	GSS	03-JUN-1999
ACCESSION	Drosophila melanogaster genome survey sequence Tm3 end of BAC #				
VERSION	BA01916 of RPCT-98 library from Drosophila melanogaster (fruit				
KEYWORDS	fly), genomic survey sequence.				
SOURCE	AL056652				
ORGANISM	AL056652.1	GI:4932342			
REFERENCE	GSS.				
AUTHORS	fruit fly.				
TITLE	Drosophila melanogaster				
JOURNAL	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
COMMENT	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;				
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
	1 (bases 1 to 844)				
	Genoscope.				
	Direct Submission				
	Submitted (02-JUN-1999) Genoscope - Centre National de Sequenc				
	BP 191 91006 EVRY cedex - FRANCE (E-mail : secrete@genoscope.cns				
	- web : www.genoscope.cns.fr)				
	Determination of this BAC end sequence was carried out as part o				

Aaron Memmoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDP from the isogenic strain y2: cn bw sp, the same strain used for the BDP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

Location/Qualifiers  
1. 844

BASE COUNT	261 a	112 c	92 g	35 t	344 others
ORIGIN					

Query Match	3.78;	Score 45.4;	DB 82;	Length 844;
Best Local Similarity	16.08;	Pred. NO. 12;		
Matches 54;	Conservative 144;	Mismatches 140;	Indels 0;	Gaps 0;

OY	886	AAGGCTACACACGACCGACCTTACGACACCGACAGTCTCCCGTGCAGCAGACAGCTGTTT	945
Db	733	MMVTMCMMMAVAAVAVYVYAMAMARNNNVYCSBCNBBTTTYCNVTCYSSSSSSBITTCM	674
OY	946	ACTGCGTGGCGACATATCCCGCCTCGGTGCGCAAGACGTCACCTACGCGGTACCAACG	1005
Db	673	THTNSNTYTWYTVTTTHATNTTCCYCCOCCCCYCSOCBIBCNITSCSYYSBSCSCS	614
OY	1006	CGACGATGCGGTATGTCCTCGACTTCGACCATGCGACCATGGTGGCGGGCGGTGCTCGTGA	1065
Db	613	SSSSSSSCSGSBSYSCCCBCCGCGYCGCYSBVCSBSSTBSYSGBSSTSSGTCKCKBS	554
OY	1066	ATCCGGCAGCTAGCCAAAGTCTTTTCGACGACGACGAAGGGCGTTACTGTCCGAACGGTGG	1125
Db	553	SSSSBBSBSBSCCCYCSYCTCYSSCCBYCCBSGBCTBSYBSGSSGDTGBSSBSGKS	494
OY	1126	CCACGACACTGGGGGCTCCCTTCACGACGAGCCCCACAGCGTGGCGCTACACTCTGGCGC	1185
Db	493	GKSSKGGGBKSGSKGYCYCBTTCBSBSBCCSSSSSTSKKCCGGSSSTTCGSSBBGG	434
OY	1186	CTCGTTACGCCGACACACTCTGCGAGTAAGGAACCG	1223
Db	433	GSSSNTGSSSSCUGSSSTGSSSTGSSSSKYSYSSSSSSG	396

RESULT 8  
 CNS006XK  
 LOCUS CNS006XK 935 bp DNA GSS 03-JUN-1999  
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #  
 BA0R1AN09 of Rpci-98 library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 AL066051  
 ACCESSION AL066051.1 GI:4945019  
 VERSION 1  
 KEYWORDS GSS:  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 935)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sege@genoscope.cns.fr  
 COMMENT - Web : www.genoscope.cns.fr)  
 Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila  
 melanogaster genome using these BACs. for further information





QY 448 TTCCCCGATGAGCAAGTCCCGGTCGACTCCGCCCAAGTTGGCGCTTACCACCTGA 507  
 Db 248 AAMCMSCMSVGGACACAGGCGGAGAACCCGCGCACSCCCSSSCSCSCSSSSS 189  
 QY 508 TCGCAACCCCAAGGGGCGCGGTGTGCTGATGAGGCGGCGGTGCCGCTGACACCGGCG 567  
 Db 188 CGGAGACCCCSAGAGGGCGCGCGCCGACCCGCGSAGACCCGAC -CMNVGGSS 130  
 QY 568 ACCTGCTGTGATCGCGCGCGGACCGCGCTACACGACCGCGCGTACCGACGGA 627  
 Db 129 ACCCMAAGGGAGAGCGCGGAGCCCGCCGACCCCGCCCGCCGACCGCGGGA 70  
 QY 628 TGGGCGGACC 638  
 Db 69 AGCGCCCAAC 59  
 RESULT 13  
 CENS00720 932 bp DNA GSS 03-JUN-1999  
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #  
 DEFINITION BAC14B09 of RPCI-98 library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 ACCESSION AL066742.1 GI:4945205  
 VERSION AL066742  
 KEYWORDS GSS.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 932)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 COMMENT Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila  
 melanogaster genome using these BACs. For further information  
 please see <http://www.fruitfly.org> The BDGP Drosophila  
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
 Aaron Mamoser in Pletier de Jong's laboratory in the Department of  
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
 NY. The library is named RPCI-98 and was constructed by partial  
 EcoRI digestion of Drosophila DNA provided by the BDGP from the  
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's  
 P1 and EST libraries. A more detailed description of the library  
 and how to order individual BAC clones, the entire library, or  
 filters for hybridization from the BACPAC Resource Center can be  
 found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
 FEATURES  
 source  
 1..932  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="RPCI-98"  
 /clone="BAC14B09"  
 /note="end : 77"  
 BASE COUNT 155 a 202 c 241 g 91 t 243 others  
 ORIGIN  
 Query Match 3.6%: Score 44.8; DB 82; Length 932;  
 Best Local Similarity 34.1%: Pred. No. 15;  
 Matches 73; Conservative 47; Mismatches 94; Indels 0; Gaps 0;  
 QY 432 GAGCGACTACCCCTGCTCCCGGATGAGCGAGGAGCGGCTGAGTCCGCCGCGGTT 491  
 Db 679 GCSGCCG 738  
 QY 492 GCGGCTTACCACTGATGCAACCAAGGCGCGCGGTGTGCTGATGGCGGGGTGCC 551

Db 739 CCSCCCSSCGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 798  
 QY 552 GAGCTGCAACCG 611  
 Db 799 SSGSGCGSSSSSSCGGSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 858  
 QY 612 CGCATCGCCACGCGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 645  
 Db 859 CGSGGCGSGCG 892  
 RESULT 14  
 CENS006XK 9 bp DNA GSS 03-JUN-1999  
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #  
 DEFINITION BAC14N09 of RPCI-98 library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 ACCESSION AL066051.1 GI:4945019  
 VERSION AL066051  
 KEYWORDS GSS.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 935)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 COMMENT Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila  
 melanogaster genome using these BACs. For further information  
 please see <http://www.fruitfly.org> The BDGP Drosophila  
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
 Aaron Mamoser in Pletier de Jong's laboratory in the Department of  
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
 NY. The library is named RPCI-98 and was constructed by partial  
 EcoRI digestion of Drosophila DNA provided by the BDGP from the  
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's  
 P1 and EST libraries. A more detailed description of the library  
 and how to order individual BAC clones, the entire library, or  
 filters for hybridization from the BACPAC Resource Center can be  
 found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
 FEATURES  
 source  
 1..935  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="RPCI-98"  
 /clone="BAC14N09"  
 /note="end : 77"  
 BASE COUNT 257 a 170 c 152 g 96 t 250 others  
 ORIGIN  
 Query Match 3.6%: Score 44.2; DB 82; Length 935;  
 Best Local Similarity 27.4%: Pred. No. 20;  
 Matches 93; Conservative 83; Mismatches 163; Indels 0; Gaps 0;  
 QY 406 CTTACGAGACCGTCCAGACCGCGGAGGCGGCTGCTGCTGCGGATGAGCGAAG 465  
 Db 856 CCGCGCGGSSGCG 797  
 QY 466 TCGCGGTGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 525  
 Db 796 SGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 737  
 QY 526 GCGGTGCTGATGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 585  
 Db 736 GCSGSSGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 677



GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:55:19 ; Search time 8627.09 Seconds  
(without alignments)  
-139.259 Million cell updates/sec

Title: US-09-362-485-8  
Sequence: 1235  
1 ATCTTCGAGATTATTCGAC.....GGAGAGCGATGATCGGCC 1235

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 862769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pl1: \*  
8: gb\_pl2: \*  
9: gb\_pl3: \*  
10: gb\_pl4: \*  
11: gb\_pl5: \*  
12: gb\_ro: \*  
13: gb\_sts: \*  
14: gb\_sy: \*  
15: gb\_un: \*  
16: gb\_vl: \*  
17: em\_fun1: \*  
18: em\_fun2: \*  
19: em\_fun3: \*  
20: em\_fun4: \*  
21: em\_fun5: \*  
22: em\_ov: \*  
23: em\_ov: \*  
24: em\_pat: \*  
25: em\_ph: \*  
26: em\_pl: \*  
27: em\_ro: \*  
28: em\_sts: \*  
29: em\_sy: \*  
30: em\_un: \*  
31: em\_vl: \*  
32: gb\_hlg1: \*  
33: gb\_hlg2: \*  
34: gb\_hlg3: \*  
35: gb\_hlg4: \*  
36: gb\_hlg5: \*  
37: gb\_hlg6: \*  
38: em\_hum1: \*  
39: em\_hum2: \*  
40: em\_hum3: \*  
41: em\_hum4: \*  
42: em\_hum5: \*  
43: em\_hum6: \*  
44: em\_hum7: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1235	100.0	1235	5 A87611	A87611 Sequence 8
2	1235	100.0	1235	5 A89751	A89751 Sequence 8
3	1232.4	99.8	1235	5 A87609	A87609 Sequence 6
4	1232.4	99.8	1235	5 A89749	A89749 Sequence 6
5	1228	99.4	1229	5 A87610	A87610 Sequence 7
6	1228	99.4	1229	5 A89750	A89750 Sequence 7
7	1222.4	99.0	1236	5 A87613	A87613 Sequence 10
8	1222.4	99.0	1236	5 A89753	A89753 Sequence 10
9	1220.8	98.9	1237	5 A87607	A87607 Sequence 4
10	1220.8	98.9	1237	5 A89747	A89747 Sequence 4
11	1220.8	98.9	56414	1 MTW002	MTW002 Sequence 4
12	1218.4	98.7	1235	5 A87606	A87606 Sequence 3
13	1218.4	98.7	1235	5 A89746	A89746 Sequence 3
14	1212.8	98.2	1228	5 A87608	A87608 Sequence 3
15	1212.8	98.2	1228	5 A89748	A89748 Sequence 3
16	1206.4	97.7	1245	5 A87605	A87605 Sequence 2
17	1206.4	97.7	1245	5 A89745	A89745 Sequence 2
18	1206.4	97.7	1260	5 A89744	A89744 Sequence 1
19	1206.4	97.7	1260	5 A89744	A89744 Sequence 1
20	1196.4	96.7	1209	1 MTALADH	MTALADH Sequence 1
21	1194.4	96.1	1209	5 A87612	A87612 Sequence 9
22	1187.4	95.6	1194	5 A89752	A89752 Sequence 9
23	1180.4	95.6	1194	2 MTU92472	MTU92472 Sequence 9
24	684.4	55.4	720	5 A89759	A89759 Sequence 16
25	524.8	42.5	40745	1 SCISL	SCISL Sequence 16
26	454	36.8	682	5 A89758	A89758 Sequence 15
27	454	36.8	682	5 A89758	A89758 Sequence 15
28	363.4	29.4	8046	2 AF049107	AF049107 Sequence 19
29	339.2	27.5	11514	2 AE002028	AE002028 Sequence 19
30	318.8	25.8	1585	2 AF070745	AF070745 Sequence 19
31	317.8	25.7	2780	1 RLE28118	RLE28118 Sequence 19
32	300.8	24.4	1736	1 BACALDH	BACALDH Sequence 19
33	279.4	22.6	1477	2 AF070715	AF070715 Sequence 19
34	276.4	22.4	2113	1 D37807	D37807 Sequence 19
35	265.2	21.5	1125	5 E01847	E01847 Sequence 19
36	249.2	20.2	137740	1 D90960	D90960 Sequence 19
37	232.8	18.9	1375	1 AB013821	AB013821 Sequence 19
38	231	18.7	1776	1 BACALDH	BACALDH Sequence 19
39	223.6	18.1	1709	1 BACALDH	BACALDH Sequence 19
40	223.6	18.1	12316	1 BS282015	BS282015 Sequence 19
41	223.6	18.1	217420	1 BSUB0017	BSUB0017 Sequence 19
42	206.2	16.7	1714	2 AF070714	AF070714 Sequence 19
43	101.6	8.2	10652	2 AE001565	AE001565 Sequence 19
44	93.6	7.6	10512	2 AE000639	AE000639 Sequence 19
45	74	6.0	3160	1 RR05294	RR05294 Sequence 19

#### ALIGNMENTS



```

RESULT 1
LOCUS      A87611      1235 bp      DNA
DEFINITION Sequence 8 from Patent WO9836089.
ACCESSION  A87611
VERSION     A87611.1 GI:6736251
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 1235)
AUTHORS     Flohe, L. and Singh, M.
TITLE       TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL     Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
            source
            1..1235
            /organism="unidentified"
            /db_xref="taxon:32644"
BASE COUNT  236 a      394 c      385 g      220 t
ORIGIN
Query Match      100.0%; Score 1235; DB 5; Length 1235;
Best Local Similarity 100.0%; Pred. No. 2,7e-168;
Matches 1235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATCTTGCAATTAATGAACTTTCTTCACTGAAAGCTATCGAGAGGGTAAATC 60
DB      1 ATCTTGCAATTAATGAACTTTCTTCACTGAAAGCTATCGAGAGGGTAAATC 60
QY      61 ATGCGCGTGGTATTCGACCGAGACAAACAAAGCAATTCGGGTGCGCATCAACCCG 120
DB      61 ATGCGCGTGGTATTCGACCGAGACAAACAAAGCAATTCGGGTGCGCATCAACCCG 120
QY      121 GCGCGCTGCGGAACTAAACCGCTGCGCATGAGTCTCATCCAGCAGTGGCGGA 180
DB      121 GCGCGCTGCGGAACTAAACCGCTGCGCATGAGTCTCATCCAGCAGTGGCGGA 180
QY      122 GCGCGCTGCGGAACTAAACCGCTGCGCATGAGTCTCATCCAGCAGTGGCGGA 180
DB      122 GCGCGCTGCGGAACTAAACCGCTGCGCATGAGTCTCATCCAGCAGTGGCGGA 180
QY      181 GAGGGCTGGGTATACGACGAGGGGATTTCAAGGGGCGAGGGCGCAATGGTGGCGAC 240
DB      181 GAGGGCTGGGTATACGACGAGGGGATTTCAAGGGGCGAGGGCGCAATGGTGGCGAC 240
QY      241 GCGGACAGAGTGTGGGCGGACGCTGATTTATGCTCAAGGTCAAGAACCGATAGCGGCG 300
DB      241 GCGGACAGAGTGTGGGCGGACGCTGATTTATGCTCAAGGTCAAGAACCGATAGCGGCG 300
QY      301 GAATACGGCGCGCTGCGACACGGGCGATCTTGTTCACGTTCTTTCGCGCGCTCAC 360
DB      301 GAATACGGCGCGCTGCGACACGGGCGATCTTGTTCACGTTCTTTCGCGCGCTCAC 360
QY      361 GTCCTTGACACGATGCTTGTGATTCGGGACACGATTCGCTTACGAGACCGTCC 420
DB      361 GTCCTTGACACGATGCTTGTGATTCGGGACACGATTCGCTTACGAGACCGTCC 420
QY      421 AGACCGCGCCTAAGGCGACATACCCCTGCTGCGCGCATATGAGCAAGTC 480
DB      421 AGACCGCGCCTAAGGCGACATACCCCTGCTGCGCGCATATGAGCAAGTC 480
QY      481 CGGCGCAAGTGTGGGCTTACACATGATCGGAACCCAAAGGGGCGCGGTGTGATG 540
DB      481 CGGCGCAAGTGTGGGCTTACACATGATCGGAACCCAAAGGGGCGCGGTGTGATG 540
QY      541 GCGGGGTGCGCGGCTGCAACCGGCGAGCGTGTGATCGGCGCGCGCGCGCGGCT 600
DB      541 GCGGGGTGCGCGGCTGCAACCGGCGAGCGTGTGATCGGCGCGCGCGCGGCT 600
QY      601 ACAACGCGACCGCGCATGCGCAAGGGGATGGGCGGCGCGCTTACAGCATCAACA 660
DB      601 ACAACGCGACCGCGCATGCGCAAGGGGATGGGCGGCGCGCTTACAGCATCAACA 660
QY      661 TCGACAAACTTGGGCACTGACGCGGAGTTCTGGCGCGCGCTTCACACTCGCTACTCAT 720

```

```

DB      661 TCGACAAACTTGGGCACTGACGCGGAGTTCTGGCGCGCGCTTCACACTCGCTACTCAT 720
QY      721 CGGCGTACGAGCTCGAGGCTGCGCTCAAGGTGCGGACCTGTGTATTTGGGCGGCTCTGG 780
DB      721 CGGCGTACGAGCTCGAGGCTGCGCTCAAGGTGCGGACCTGTGTATTTGGGCGGCTCTGG 780
QY      781 TGGCAGGCGCGCAAGGCAACCAATTAATCTTCTGAAATTAATCTTGTGCGCATATGAAACGAG 840
DB      781 TGGCAGGCGCGCAAGGCAACCAATTAATCTTCTGAAATTAATCTTGTGCGCATATGAAACGAG 840
QY      841 GTGGGTAATGCTGGTATATACCATTCGACGAGGCGGCTGTTTGAAGGCTCAAGACGGA 900
DB      841 GTGGGTAATGCTGGTATATACCATTCGACGAGGCGGCTGTTTGAAGGCTCAAGACGGA 900
QY      901 CCACCTACGACCCACCCGACGTTGCGCGGTGACGACACGCTGTTTACTGCTGGGCGAACA 960
DB      901 CCACCTACGACCCACCCGACGTTGCGCGGTGACGACACGCTGTTTACTGCTGGGCGAACA 960
QY      961 TGCCCGGCTGGTGGCGGACGCTGACCTTACGCGGTGACCAACGCGCATGCGCGTATG 1020
DB      961 TGCCCGGCTGGTGGCGGACGCTGACCTTACGCGGTGACCAACGCGCATGCGCGTATG 1020
QY      1021 TGCTGACCTTGGCGGACCATGCTGCGGCGGCGGCTGCGGATCGGATCGGCACTAGCA 1080
DB      1021 TGCTGACCTTGGCGGACCATGCTGCGGCGGCGGCTGCGGATCGGATCGGCACTAGCA 1080
QY      1081 AAGGCTCTTTCACGACGACGAGAGGCGCTTACTGTCCGAAGGCTGCGGCACTGCGGCG 1140
DB      1081 AAGGCTCTTTCACGACGACGAGAGGCGCTTACTGTCCGAAGGCTGCGGCACTGCGGCG 1140
QY      1141 TGCCGCTACGACGCGCGCGAGCTGCTGCGGCTGACCTCGCGCGCTGTACCGCAGC 1200
DB      1141 TGCCGCTACGACGCGCGCGAGCTGCTGCGGCTGACCTCGCGCGCTGTACCGCAGC 1200
QY      1201 ACAGCTGGGAGTAAGGAGACGATGATGTGCGCC 1235
DB      1201 ACAGCTGGGAGTAAGGAGACGATGATGTGCGCC 1235

RESULT 2
LOCUS      A89751      1235 bp      DNA
DEFINITION Sequence 8 from Patent WO9832862.
ACCESSION  A89751
VERSION     A89751.1 GI:6738285
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 1235)
AUTHORS     Flohe, L. and Singh, M.
TITLE       L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL     Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
            source
            1..1235
            /organism="unidentified"
            /db_xref="taxon:32644"
BASE COUNT  236 a      394 c      385 g      220 t
ORIGIN
Query Match      100.0%; Score 1235; DB 5; Length 1235;
Best Local Similarity 100.0%; Pred. No. 2,7e-168;
Matches 1235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATCTTGCAATTAATGAACTTTCTTCACTGAAAGCTATCGAGAGGGTAAATC 60
DB      1 ATCTTGCAATTAATGAACTTTCTTCACTGAAAGCTATCGAGAGGGTAAATC 60
QY      61 ATGCGCGTGGTATTCGACCGAGACAAACAAAGCAATTCGGGTGCGCATCAACCCG 120

```

D	b	61	ATGCGCTGCGGTATTTCGACACGAGAACCAAAAACAAGAAATTCGGGTGGCCATCACCCG	120
O	y	121	GCCGGCGTCGCGGAACCTAACCCGCTGTCGCCATAGATGCTCATCCAGCGAGTCCGGA	180
D	b	121	GCCGCGCTGCGGGAACCTAACCCGCTGTCGCCATAGATGCTCATCCAGCGAGTCCGGA	180
O	y	181	GAGGGCTGCGGTATCCACGACGGGAAATTTAAAGGCGGAGGGCGCAATGGTGGCGAC	240
D	b	181	GAGGGCTGCGGTATCCACGACGGGAAATTTAAAGGCGGAGGGCGCAATGGTGGCGAC	240
O	y	241	GCCGCAAGGTGTGGGGCGACGCTGATTTATTCCTCAAGTCAAAAACCGATACGGCG	300
D	b	241	GCCGCAAGGTGTGGGGCGACGCTGATTTATTCCTCAAGTCAAAAACCGATACGGCG	300
O	y	301	GAATACGGGCCCCCTGCGACACAGGGCCATCTGTTCACGTTCTTCGATTGGCCGCTAC	360
D	b	301	GAATACGGGCCCCCTGCGACACAGGGCCATCTGTTCACGTTCTTCGATTGGCCGCTAC	360
O	y	361	GTCGTTGACACGATGACGTTGTGTGGATTCCGCGACACAGTCAATGGCTACAGACGTC	420
D	b	361	GTCGTTGACACGATGACGTTGTGTGGATTCCGCGACACAGTCAATGGCTACAGACGTC	420
O	y	421	AGACCGCCGAAGGCGCACTACCCCTGCTTGCCTCCCGATGACCGAAGTGCCTGATCG	480
D	b	421	AGACCGCCGAAGGCGCACTACCCCTGCTTGCCTCCCGATGACCGAAGTGCCTGATCG	480
O	y	481	CGGCCCAGTGTGGGCGTTTACACCTGATGCGAACCCAAAGGGGCGCGGTGTGATAGG	540
D	b	481	CGGCCCAGTGTGGGCGTTTACACCTGATGCGAACCCAAAGGGGCGCGGTGTGATAGG	540
O	y	541	GCGGGGTGCCCCGCGCTGCAACCGGCGCAGCTCGTGTGATCGGCGCGGCAACCGCGCT	600
D	b	541	GCGGGGTGCCCCGCGCTGCAACCGGCGCAGCTCGTGTGATCGGCGCGGCAACCGCGCT	600
O	y	601	ACAAGCGAGCCCCGATGCGCAAGGGCATGGGGCCGACCGTTACGGTTCTAAGCATCAACA	660
D	b	601	ACAAGCGAGCCCCGATGCGCAAGGGCATGGGGCCGACCGTTACGGTTCTAAGCATCAACA	660
O	y	661	TGCAACAACTTCGCGCACTGCGACGCGCAGATTTCGCGCGGATTCACACTGCTACTCAT	720
D	b	661	TGCAACAACTTCGCGCACTGCGACGCGCAGATTTCGCGCGGATTCACACTGCTACTCAT	720
O	y	721	CGGCGTACGAGCTCGAAGGTGCGTCAAAAGTGCACCTGGTATGGGCGCTCTGG	780
D	b	721	CGGCGTACGAGCTCGAAGGTGCGTCAAAAGTGCACCTGGTATGGGCGCTCTGG	780
O	y	781	TGCCAGGGGCCAAGCGACCCCAATTATGTCTGCAATTCACCTGTGCGGCAATTAAGAACAG	840
D	b	781	TGCCAGGGGCCAAGCGACCCCAATTATGTCTGCAATTCACCTGTGCGGCAATTAAGAACAG	840
O	y	841	GTCGGTACTGTGTGATATAGCCATGACACAGGGCGGCTGTTCGAAGGCTACACACGA	900
D	b	841	GTCGGTACTGTGTGATATAGCCATGACACAGGGCGGCTGTTCGAAGGCTACACACGA	900
O	y	901	CCACCTACGACACACCGACGTTGGCGGTGACACACGCTGTTTACGTGCGTGGGACACA	960
D	b	901	CCACCTACGACACACCGACGTTGGCGGTGACACACGCTGTTTACGTGCGTGGGACACA	960
O	y	961	TGCCCGGCTCGGTGCGAAGACGTGCACTACGCGCTGACCAACGGGAGATGCGGTATG	1020
D	b	961	TGCCCGGCTCGGTGCGAAGACGTGCACTACGCGCTGACCAACGGGAGATGCGGTATG	1020
O	y	1021	TGCTCGAGCTTGGCGACATAGGCTGGGCGGGCGCGTGGCGGTGCAATCCGGCATACGCA	1080
D	b	1021	TGCTCGAGCTTGGCGACATAGGCTGGGCGGGCGCGTGGCGGTGCAATCCGGCATACGCA	1080
O	y	1081	AAGGTCCTTCGACGACAGAAAGGGCGTTACTGTCCGAACGGGTGGGCAACGACCTGGGGG	1140
D	b	1081	AAGGTCCTTCGACGACAGAAAGGGCGTTACTGTCCGAACGGGTGGGCAACGACCTGGGGG	1140
O	y	1141	TGCGGTTACGAGACCCGCGCAAGGCTGTGCGCTGACATCTTCGCGCGCTGTACGCGGAC	1200
D	b	1141	TGCGGTTACGAGACCCGCGCAAGGCTGTGCGCTGACATCTTCGCGCGCTGTACGCGGAC	1200

QY	1201	ACAGTCGGGAGTAAGAAGGAAGCAGTGCGCC	1235
Db	1201	ACAGTCGGGAGTAAGAAGGAAGCAGTGCGCC	1235
RESULT	3		
A87609			
LOCUS	A87609	1235 bp	DNA
DEFINITION	Sequence 6 from Patent WO936089.		PAT
ACCESSION	A87609		
VERSION	A87609.1	GI:6736249	
KEYWORDS	.		
SOURCE	unidentified.		
ORGANISM	unidentified		
REFERENCE	unclassified.		
AUTHORS	1 (bases 1 to 1235)		
TITLE	Flohe,L. and Singh,M.		
JOURNAL	TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE		
	Patent: WO 9836089-A 20-AUG-1998;		
	FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)		
FEATURES	Location/Qualifiers		
Source	1..1235		
	/organism="unidentified"		
	/db_xref="taxon:32644"		
BASE COUNT	235 a 395 c 384 g 220 t		1 others
ORIGIN			

Query Match	99.8%;	Score 1232.4;	DB 5;	Length 1235;
Best Local Similarity	99.8%;	Pred. No. 6.3e-168;		
Matches 1233;	Conservative	0;	Mismatches 2;	Indels 0;
				Gaps 0;

QY	1	ACCTCGAGATTAATACGAACTTTCCACACTGAAAGCTACAGTATCGAGGGGTATC	60
Db	1	ACTCTGCAGATTAAATCGAACTTTCTTCACACTGAAAGCTACAGTATCGAGGGGTATC	60
QY	61	ATGCGCGTCGTGATTCCGACCGAGACCAAAAACAAGTAATCCGGGTGGCCATACCCG	120
Db	61	ATGGCGCGTCGTGATTCCGACCGAGACCAAAAACAAGTAATCCGGGTGGCCATACCCG	120
QY	121	GCCGGCGTCGCGGAAGTAACCCGTCGTGGCCATGAGTGTCTATCCAGCAGGTGCCGA	180
Db	121	GCCGCGCGTCGCGGAAGTAACCGTCGTGGCCATGAGTGTCTATCCAGCAGGTGCCGA	180
QY	181	GAGGGCTGCGGCTATCCAGCACCCGATTCACAGCGCGAGCCCGCAACTGTCGGCAC	240
Db	181	GAGGGCTGCGGCTATCCAGCACCGGATTTCAAGGCGCGAGCGCGCAACTGTCGGCAC	240
QY	241	GCCGACCAAGGTGTGGGGCGAGCGGTGATTTATGTGTCAAGGTCAAAAGAACGATAGCGCG	300
Db	241	GCCGACCAAGGTGTGGGGCGAGCGGTGATTTATGTGTCAAGGTCAAAAGAACGATAGCGCG	300
QY	301	GAAATACGCGCGCTGCCAGACGGGGGAGTCTTTGACGCTTTGCAATTTGGCCGCTAC	360
Db	301	GAAATACGCGCGCTGCCAGACGGGGGAGTCTTTGACGCTTTGCAATTTGGCCGCTAC	360
QY	361	GTCGTTGCACCGAGTCGTTGTTGGATTCCGGCACCGTCATATTGCCACGAGACCGTCC	420
Db	361	GTCGTTGCACCGAGTCGTTGTTGGATTCCGGCACCGTCATATTGCCACGAGACCGTCC	420
QY	421	AGACCGCGAAGGCGCATATACCCCTGCTTGGCCCGATAGGGAAGTCCGCGGTGACTCG	480
Db	421	AGACCGCGCGAGCGCGCATATACCCCTGCTTGGCTTGGCCCGATAGGGAAGTCCGCGGTGACTCG	480
QY	481	CCGCCAGGTTGGCGCTTACCACTGATGCGAACCAAGGGGGCGCGGTGTGCTGATGG	540
Db	481	CCGCCAGGTTGGCGCTTACCACTGATGCGAACCAAGGGGGCGCGGTGTGCTGATGG	540
QY	541	GCGGGGTGCGCGGCTCTAACCAGGCGAGCTGTGTGTATCGGCGCGACCGCGGCT	600
Db	541	GCGGGGTGCGCGGCTCTAACCAGGCGAGCTGTGTGTATCTCGCGCGACCGCGGCT	600

```

QY 601 ACAACGACCCGATCGCCAAAGGATGGGCGGACCGTACGCTTACATCAACA 660
DB 601 ACAACGACCCGATCGCCAAAGGATGGGCGGACCGTACGCTTACATCAACA 660
QY 661 TCGACAACCTTGGGCACTGACGCGGAGTTCTGGGCGGATCCACTCGCTACTCAT 720
DB 661 TCGACAACCTTGGGCACTGACGCGGAGTTCTGGGCGGATCCACTCGCTACTCAT 720
QY 721 CGGCGTACGAGCTCGAGGGTGGCGCTCAAAAGTGGCGGATGGGCGGCTGCTGG 780
DB 721 CGGCGTACGAGCTCGAGGGTGGCGCTCAAAAGTGGCGGATGGGCGGCTGCTGG 780
QY 781 TCGCAGGCGCCAAAGGACCCCAATTAAGTCTGCAATTCATTGCGGATATGAACAG 840
DB 781 TCGCAGGCGCCAAAGGACCCCAATTAAGTCTGCAATTCATTGCGGATATGAACAG 840
QY 841 GTGCGGTACTGTGATATATAGCATGACAGGCGGCTGTTTCAAGGCTACAGCGGA 900
DB 841 GTGCGGTACTGTGATATATAGCATGACAGGCGGCTGTTTCAAGGCTACAGCGGA 900
QY 901 CCACCTACGACACCCGACGTTGCGGCTGACAGACGCTGTTTACTGCGTGGGAGCA 960
DB 901 CCACCTACGACACCCGACGTTGCGGCTGACAGACGCTGTTTACTGCGTGGGAGCA 960
QY 961 TCGCCGCTCGGTGCGGAAAGAGCTGACCTAGCGCTGACCAACGCGAGATGCCGTATG 1020
DB 961 TCGCCGCTCGGTGCGGAAAGAGCTGACCTAGCGCTGACCAACGCGAGATGCCGTATG 1020
QY 1021 TCGCTGAGCTTGGCGCATGCTGCGGCGGCGGCGGCTGCGGATCCGGCATAGCCA 1080
DB 1021 TCGCTGAGCTTGGCGCATGCTGCGGCGGCGGCGGCTGCGGATCCGGCATAGCCA 1080
QY 1081 AAGGCTTTTCAGACGACGAAGGGGCGTTACTGTCCGAAAGGCTGCGGAGCTGGGG 1140
DB 1081 AAGGCTTTTCAGACGACGAAGGGGCGTTACTGTCCGAAAGGCTGCGGAGCTGGGG 1140
QY 1141 TCGCGTTTACCGAGCCCGGCAAGCTGCTGAGCTTCTGCGGCGCTGTTACGCCGAGC 1200
DB 1141 TCGCGTTTACCGAGCCCGGCAAGCTGCTGAGCTTCTGCGGCGCTGTTACGCCGAGC 1200
QY 1201 ACACGTCGGGAGTAGGGAAGCGATGTCGGGCC 1235
DB 1201 ACACGTCGGGAGTAGGGAAGCGATGTCGGGCC 1235

RESULT 4
LOCUS A89749 1235 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 6 from Patent WO9832862.
ACCESSION A89749
VERSION A89749.1 GI:6738283
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1235)
AUTHORS Flohe, U. and Singh, M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
SOURCE
1. location/Qualifiers
/db_xref="taxon:32644"
BASE COUNT 235 a 395 c 384 g 220 t 1 others
ORIGIN
Query Match 99.8%; Score 1232.4; DB 5; Length 1235;
Best Local Similarity 99.8%; Pred. No. 6.3e-168;
Matches 1233; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AACTTGACATTAATGACACTTCTTCACACGTAAGCGTACAGTATCGAGAGGGTAATC 60

```

```

DB 1 AACTTGACATTAATGACACTTCTTCACACGTAAGCGTACAGTATCGAGAGGGTAATC 60
QY 61 ATGGCGCTCGGTATTCGACCGAGACCAAAACAAAGATTCGCGGTGGCATACCCCG 120
DB 61 ATGGCGCTCGGTATTCGACCGAGACCAAAACAAAGATTCGCGGTGGCATACCCCG 120
QY 121 GCGGCGCTCGGGAATCAACCCCTGCTGAGGCTAGGCTGCTATCCAGGAGGTGCCGA 180
DB 121 GCGGCGCTCGGGAATCAACCCCTGCTGAGGCTAGGCTGCTATCCAGGAGGTGCCGA 180
QY 181 GAGGCGCTCGGATATACCGAGCGGATTCGAAGGGGCGAGCGGCACTGCTGGGACC 240
DB 181 GAGGCGCTCGGATATACCGAGCGGATTCGAAGGGGCGAGCGGCACTGCTGGGACC 240
QY 241 GCGGACCAAGTGTGGGCGGACCTGATTTATTTCTCAAGGTCAAGAAACGATAGCGCG 300
DB 241 GCGGACCAAGTGTGGGCGGACCTGATTTATTTCTCAAGGTCAAGAAACGATAGCGCG 300
QY 301 GAATACGGCGCGCTGCGACAGGGGATGCTGTCACGTTCTTGATTTGGCGCGCTAC 360
DB 301 GAATACGGCGCGCTGCGACAGGGGATGCTGTCACGTTCTTGATTTGGCGCGCTAC 360
QY 361 GTGCTTGCACCGATGCGTTGTTGATTCGGGACACGCTCAATTGCTACAGACCGCTCC 420
DB 361 GTGCTTGCACCGATGCGTTGTTGATTCGGGACACGCTCAATTGCTACAGACCGCTCC 420
QY 421 AGACGCGCGGAAGCGGACCTACCCCTGCTGCGCGGATGAGGAAGTCCCGCTGACTCG 480
DB 421 AGACGCGCGGAAGCGGACCTACCCCTGCTGCGCGGATGAGGAAGTCCCGCTGACTCG 480
QY 481 CCGCCCAAGTTGGCGCTTACCACTGATGACCAACCAAGGGGCGGCTGTGCTGATGG 540
DB 481 CCGCCCAAGTTGGCGCTTACCACTGATGACCAACCAAGGGGCGGCTGTGCTGATGG 540
QY 541 GCGGGGTGCGGCGCTG...ACGGGCGGACGCTGCTGCTGATGCA GCGCGGACCGCGGCT 600
DB 541 GCGGGGTGCGGCGCTG...ACGGGCGGACGCTGCTGCTGATGCA GCGCGGACCGCGGCT 600
QY 601 ACAACGACCCGCGCATTCGCCCAAGGACATGGGCGGAGCTTACGTTTACATCAACA 660
DB 601 ACAACGACCCGCGCATTCGCCCAAGGACATGGGCGGAGCTTACGTTTACATCAACA 660
QY 661 TCGACAACCTTGGGCACTGACGCGGAGTTCTGGGCGGATCCACTCGCTACTCAT 720
DB 661 TCGACAACCTTGGGCACTGACGCGGAGTTCTGGGCGGATCCACTCGCTACTCAT 720
QY 721 CGGCGTACGAGCTCGAGGGTGGCGCTCAAAAGTGGCGGATGGGCGGCTGCTGG 780
DB 721 CGGCGTACGAGCTCGAGGGTGGCGCTCAAAAGTGGCGGATGGGCGGCTGCTGG 780
QY 781 TCGCAGGCGCCAAAGGACCCCAATTAAGTCTGCAATTCATTGCGGATATGAACAG 840
DB 781 TCGCAGGCGCCAAAGGACCCCAATTAAGTCTGCAATTCATTGCGGATATGAACAG 840
QY 841 GTGCGGTACTGTGATATATAGCATGACAGGCGGCTGTTTCAAGGCTACAGCGGA 900
DB 841 GTGCGGTACTGTGATATATAGCATGACAGGCGGCTGTTTCAAGGCTACAGCGGA 900
QY 901 CCACCTACGACACCCGACGTTGCGGCTGACAGACGCTGTTTACTGCGTGGGAGCA 960
DB 901 CCACCTACGACACCCGACGTTGCGGCTGACAGACGCTGTTTACTGCGTGGGAGCA 960
QY 961 TCGCCGCTCGGTGCGGAAAGAGCTGACCTAGCGCTGACCAACGCGAGATGCCGTATG 1020
DB 961 TCGCCGCTCGGTGCGGAAAGAGCTGACCTAGCGCTGACCAACGCGAGATGCCGTATG 1020
QY 1021 TCGCTGAGCTTGGCGCATGCTGCGGCGGCGGCGGCTGCGGATCCGGCATAGCCA 1080
DB 1021 TCGCTGAGCTTGGCGCATGCTGCGGCGGCGGCGGCTGCGGATCCGGCATAGCCA 1080
QY 1081 AAGGCTTTTCAGACGACGAAGGGGCGTTACTGTCCGAAAGGCTGCGGACCGACTGGGGG 1140

```

```

Db 1081 AAGGTTTCGACGACGAAAGGGGCTTACTGTCCGAAAGCGGTGGCCACGACTGGGGG 1140
QY 1141 TGCGCTTACACGAGCCCGCAGCGTGTGCTGACTGTGGCGGCTGTATACCCGAGC 1200
Db 1141 TGCGCTTACACGAGCCCGCAGCGTGTGCTGACTGTGGCGGCTGTATACCCGAGC 1200
QY 1201 ACACGTCGGAGTAAGGAAGCGATGATGTCGGCC 1235
Db 1201 ACACGTCGGAGTAAGGAAGCGATGATGTCGGCC 1235

RESULT 5
LOCUS A87610 1229 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9836089.
ACCESSION A87610
VERSION A87610.1 GI:6736250
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1229)
AUTHORS Flohe, L. and Singh, M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A-20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
source 1..1229
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 236 a 391 c 382 g 219 t 1 others
ORIGIN

Query Match 99.48; Score 1228; DB 5; Length 1229;
Best Local Similarity 99.98; Freq. No. 2.7e-167;
Matches 1228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCTGAGATTATCGACTTTCTTCACTGAGCGTACGATATCGAGGGTATATC 60
Db 1 ATCTGAGATTATCGACTTTCTTCACTGAGCGTACGATATCGAGGGTATATC 60
QY 61 ATGGCGCTGGTATTCGACCGAGACCAAAACAGCAATTCGGGGTGCATCACC 120
Db 61 ATGGCGCTGGTATTCGACCGAGACCAAAACAGCAATTCGGGGTGCATCACC 120
QY 121 GCGGCGCTCGGGAATACCCGTCGTGGCATGAGGTGCTCATCCAGGAGTGC 180
Db 121 GCGGCGCTCGGGAATACCCGTCGTGGCATGAGGTGCTCATCCAGGAGTGC 180
QY 181 GAGGCTTCGCTATACCGACGCGGATTCAGAGCGGCGGCGCAACTGTCG 240
Db 181 GAGGCTTCGCTATACCGACGCGGATTCAGAGCGGCGGCGCAACTGTCG 240
QY 241 GCGGACCAAGTGTGGCGGACGCTGATTTATGCTCAAGTCAAGAACGAT 300
Db 241 GCGGACCAAGTGTGGCGGACGCTGATTTATGCTCAAGTCAAGAACGAT 300
QY 301 GAATACGGCGCTCGACACAGGCGATCTGTTCACGTTCTTGCAATTTG 360
Db 301 GAATACGGCGCTCGACACAGGCGATCTGTTCACGTTCTTGCAATTTG 360
QY 361 GTGCTTGCACGATGCTGTGTGATTCGGGACGACGTAATGCTAGAGAC 420
Db 361 GTGCTTGCACGATGCTGTGTGATTCGGGACGACGTAATGCTAGAGAC 420
QY 421 AGACGGCGAAGGCGCACTACCTGCTTGGCCGATGAGGAGTGCCTGACT 480
Db 421 AGACGGCGAAGGCGCACTACCTGCTTGGCCGATGAGGAGTGCCTGACT 480
QY 481 CCGCCAGAGTTGGCGCTTACCACTGATGCAACCAAGGGGGCGGGTGTGAT 540
Db 481 CCGCCAGAGTTGGCGCTTACCACTGATGCAACCAAGGGGGCGGGTGTGAT 540

```

```

QY 541 GCGGGGTGCCGGGCTTCGAAACCGCGCACTGCTGTGATGCGGGCGGCGCGGCT 600
Db 541 GCGGGGTGCCGGGCTTCGAAACCGCGCACTGCTGTGATGCGGGCGGCGCGGCT 600
QY 601 ACACGACGCGGCACTTCGCAACGCGATGGCGGCGGCGGCGGCGGCGGCGG 660
Db 601 ACACGACGCGGCACTTCGCAACGCGATGGCGGCGGCGGCGGCGGCGGCGG 660
QY 661 TCGACAAACTTGGCACTTCGACGCGGAGTTCTGCGGCGGATCCACTGCT 720
Db 661 TCGACAAACTTGGCACTTCGACGCGGAGTTCTGCGGCGGATCCACTGCT 720
QY 721 CGGCTACGAGCTGAG, TGCGCTCAAAAGCGCCGACCTGGTGTGAGGCG 780
Db 721 CGGCTACGAGCTGAGGAGTTCGCAAAAGCGCGGAGCTGCTGATTTGG 780
QY 781 TCGCAGGCGCCGACGCGCAATTAAGTCTCGAATTCATTTGTCGCAAT 840
Db 781 TCGCAGGCGCCGACGCGCAATTAAGTCTCGAATTCATTTGTCGCAAT 840
QY 841 TCGCGTACTGGTATATAGCATCGACGAGGCGGCGGCTTGAAGGCTCAG 900
Db 841 TCGCGTACTGGTATATAGCATCGACGAGGCGGCGGCTTGAAGGCTCAG 900
QY 901 CCACCTACGACCGCGGACGCTGCGCGTGCACGACGCTGTTTACTGCG 960
Db 901 CCACCTACGACCGCGGACGCTGCGCGTGCACGACGCTGTTTACTGCG 960
QY 961 TCGCGGCTGCTGCGGAGAGAGCTGACCTACGCGCTGACCAACCGAG 1020
Db 961 TCGCGGCTGCTGCGGAGAGAGCTGACCTACGCGCTGACCAACCGAG 1020
QY 1021 TCGTGCAGCTTCCGCACTGCTGCGGCGGCGGCGGCGGCGGCGGCG 1080
Db 1021 TCGTGCAGCTTCCGCACTGCTGCGGCGGCGGCGGCGGCGGCGGCG 1080
QY 1081 AAGGTTTCGACGCGACGAAAGGGGCTTACTGTCCGCAAGGCTGCG 1140
Db 1081 AAGGTTTCGACGCGACGAAAGGGGCTTACTGTCCGCAAGGCTGCG 1140
QY 1141 TGCGCTTACACGAGCCCGCAGCGTGTGCTGACTGTGGCGGCTGTAG 1200
Db 1141 TGCGCTTACACGAGCCCGCAGCGTGTGCTGACTGTGGCGGCTGTAG 1200
QY 1201 ACACGTCGGAGTAAGGAAGCGATGATG 1229
Db 1201 ACACGTCGGAGTAAGGAAGCGATGATG 1229

RESULT 6
LOCUS A89750 1229 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9832862.
ACCESSION A89750
VERSION A89750.1 GI:6738284
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1229)
AUTHORS Flohe, L. and Singh, M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A-30-JUN-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
source 1..1229
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 236 a 391 c 382 g 219 t 1 others
ORIGIN

```

Query Match 99.4% Score 1228; DB 5; Length 1236;  
 Best Local Similarity 99.9%; Pred. No. 2.7e-167;  
 Matches 1228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATCTTGACAGATTAAATCGAATCTTCTTCACTGAGCGGTACAGTATCGAGAGGGTAATC 60  
 1 ATCTTGACAGATTAAATCGAATCTTCTTCACTGAGCGGTACAGTATCGAGAGGGTAATC 60

61 ATGCGGTGCTGATTCGACGAGACCAAAACAAAGAAATTCGGGGTGCATCACCCG 120  
 61 ATGCGGTGCTGATTCGACGAGACCAAAACAAAGAAATTCGGGGTGCATCACCCG 120

61 ATGCGGTGCTGATTCGACGAGACCAAAACAAAGAAATTCGGGGTGCATCACCCG 120  
 61 ATGCGGTGCTGATTCGACGAGACCAAAACAAAGAAATTCGGGGTGCATCACCCG 120

121 GCGGGGTGCGGAGACTAACCGGTGCGCATGAGTCTCATCGAGCAGATGCGCGA 180  
 121 GCGGGGTGCGGAGACTAACCGGTGCGCATGAGTCTCATCGAGCAGATGCGCGA 180

121 GCGGGGTGCGGAGACTAACCGGTGCGCATGAGTCTCATCGAGCAGATGCGCGA 180  
 121 GCGGGGTGCGGAGACTAACCGGTGCGCATGAGTCTCATCGAGCAGATGCGCGA 180

181 GAGGGCTGCGCTATACACGACGCGATTTTCAAGGCGGACGCGCACTGCTGCGCAC 240  
 181 GAGGGCTGCGCTATACACGACGCGATTTTCAAGGCGGACGCGCACTGCTGCGCAC 240

181 GAGGGCTGCGCTATACACGACGCGATTTTCAAGGCGGACGCGCACTGCTGCGCAC 240  
 181 GAGGGCTGCGCTATACACGACGCGATTTTCAAGGCGGACGCGCACTGCTGCGCAC 240

241 GCGGACGAGGTGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGAACGATAGCGCG 300  
 241 GCGGACGAGGTGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGAACGATAGCGCG 300

241 GCGGACGAGGTGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGAACGATAGCGCG 300  
 241 GCGGACGAGGTGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGAACGATAGCGCG 300

301 GAATACGCGCGCTGCGACGAGCGGATCTGCTTCACTTCTGATTTGCGCGGCTCAC 360  
 301 GAATACGCGCGCTGCGACGAGCGGATCTGCTTCACTTCTGATTTGCGCGGCTCAC 360

301 GAATACGCGCGCTGCGACGAGCGGATCTGCTTCACTTCTGATTTGCGCGGCTCAC 360  
 301 GAATACGCGCGCTGCGACGAGCGGATCTGCTTCACTTCTGATTTGCGCGGCTCAC 360

361 GTGCTTGACAGCGATGCGTGTGATTCGCGGACACGATCAATGGCTACGAGACCGTCC 420  
 361 GTGCTTGACAGCGATGCGTGTGATTCGCGGACACGATCAATGGCTACGAGACCGTCC 420

361 GTGCTTGACAGCGATGCGTGTGATTCGCGGACACGATCAATGGCTACGAGACCGTCC 420  
 361 GTGCTTGACAGCGATGCGTGTGATTCGCGGACACGATCAATGGCTACGAGACCGTCC 420

421 AAGCGCGGAGGCGGACGACGCGGCTGCGCGGATGCGGAGGCGGAGGCGGCGACTG 480  
 421 AAGCGCGGAGGCGGACGACGCGGCTGCGCGGATGCGGAGGCGGAGGCGGCGACTG 480

421 AAGCGCGGAGGCGGACGACGCGGCTGCGCGGATGCGGAGGCGGAGGCGGCGACTG 480  
 421 AAGCGCGGAGGCGGACGACGCGGCTGCGCGGATGCGGAGGCGGAGGCGGCGACTG 480

481 CCGCGGAGGTGGCGCTTACCACTGATGCGAACCCAAAGGGGCGCGGCTGCTGATG 540  
 481 CCGCGGAGGTGGCGCTTACCACTGATGCGAACCCAAAGGGGCGCGGCTGCTGATG 540

481 CCGCGGAGGTGGCGCTTACCACTGATGCGAACCCAAAGGGGCGCGGCTGCTGATG 540  
 481 CCGCGGAGGTGGCGCTTACCACTGATGCGAACCCAAAGGGGCGCGGCTGCTGATG 540

541 GCGGGGTGCGGCGGCTGCGAACGCGGACGATGCTGATTCGCGCGGCGGACGCGCGCT 600  
 541 GCGGGGTGCGGCGGCTGCGAACGCGGACGATGCTGATTCGCGCGGCGGACGCGCGCT 600

541 GCGGGGTGCGGCGGCTGCGAACGCGGACGATGCTGATTCGCGCGGCGGACGCGCGCT 600  
 541 GCGGGGTGCGGCGGCTGCGAACGCGGACGATGCTGATTCGCGCGGCGGACGCGCGCT 600

601 ACAAGCGAGCGGCGATGCGCAAGCGGATGCGGCGGCGGACGCTTACGATAGACATCA 660  
 601 ACAAGCGAGCGGCGATGCGCAAGCGGATGCGGCGGCGGACGCTTACGATAGACATCA 660

601 ACAAGCGAGCGGCGATGCGCAAGCGGATGCGGCGGCGGACGCTTACGATAGACATCA 660  
 601 ACAAGCGAGCGGCGATGCGCAAGCGGATGCGGCGGCGGACGCTTACGATAGACATCA 660

661 TCGACAAACTTGGGCAACTGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 720  
 661 TCGACAAACTTGGGCAACTGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 720

661 TCGACAAACTTGGGCAACTGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 720  
 661 TCGACAAACTTGGGCAACTGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 720

721 CCGGCTTACGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 780  
 721 CCGGCTTACGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 780

721 CCGGCTTACGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 780  
 721 CCGGCTTACGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 780

781 TCGCAGGCGCGCAAGGCGCAAAATAGTCTGCAATCTGCTGCGGATGCAATGCAACAG 840  
 781 TCGCAGGCGCGCAAGGCGCAAAATAGTCTGCAATCTGCTGCGGATGCAATGCAACAG 840

781 TCGCAGGCGCGCAAGGCGCAAAATAGTCTGCAATCTGCTGCGGATGCAATGCAACAG 840  
 781 TCGCAGGCGCGCAAGGCGCAAAATAGTCTGCAATCTGCTGCGGATGCAATGCAACAG 840

841 GTGCGGTGCTGATGATAGCCATGACGAGGCGGCTGTTGCAAGGCTCACAGCGCA 900  
 841 GTGCGGTGCTGATGATAGCCATGACGAGGCGGCTGTTGCAAGGCTCACAGCGCA 900

841 GTGCGGTGCTGATGATAGCCATGACGAGGCGGCTGTTGCAAGGCTCACAGCGCA 900  
 841 GTGCGGTGCTGATGATAGCCATGACGAGGCGGCTGTTGCAAGGCTCACAGCGCA 900

901 CCGCCTTACGAGCGGCGGATGCGGCGGATGCGGCGGCGGATGCACTGCTGATCAT 960  
 901 CCGCCTTACGAGCGGCGGATGCGGCGGATGCGGCGGCGGATGCACTGCTGATCAT 960

901 CCGCCTTACGAGCGGCGGATGCGGCGGATGCGGCGGCGGATGCACTGCTGATCAT 960  
 901 CCGCCTTACGAGCGGCGGATGCGGCGGATGCGGCGGCGGATGCACTGCTGATCAT 960

961 TCGCAGGCGCGCAAGGCGCAAGGCGGATGCGGCGGATGCGGCGGATGCACTGCTGATCAT 1020  
 961 TCGCAGGCGCGCAAGGCGCAAGGCGGATGCGGCGGATGCGGCGGATGCACTGCTGATCAT 1020

961 TCGCAGGCGCGCAAGGCGCAAGGCGGATGCGGCGGATGCGGCGGATGCACTGCTGATCAT 1020  
 961 TCGCAGGCGCGCAAGGCGCAAGGCGGATGCGGCGGATGCGGCGGATGCACTGCTGATCAT 1020

1021 TCGCTGAGCTTGGCGACATGCTGCGGCGGCGGCTGCTGCTGCAATCCGCGCACTAGCCA 1080  
 1021 TCGCTGAGCTTGGCGACATGCTGCGGCGGCGGCTGCTGCTGCAATCCGCGCACTAGCCA 1080

1236 bp DNA PAT 22-JAN-2000

LOCUS A87613 1236 bp DNA PAT 22-JAN-2000

DEFINITION Sequence 10 from Patent WO9836089.

ACCESSION A87613

VERSION A87613.1 GI:6736253

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 1236)

AUTHORS

Flohe, L. and Singh, M.

TITLE

TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE

JOURNAL

PATENT: WO 9836089-A 20-AUG-1998;

FLOHE LEOBOLD (DE); SINGH MAHAVIR (DE)

FEATURES

Location/Qualifiers

1..1236

source

1..1236

BASE COUNT 236 a 395 c 385 g 220 t

ORIGIN

1..1236

Query Match 99.0% Score 1222.4; DB 5; Length 1236;  
 Best Local Similarity 99.8%; Pred. No. 1.7e-166;  
 Matches 1234; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

1 ATCTTGACAGATTAAATCGAATCTTCTTCACTGAGCGGTACAGTATCGAGAGGGTAATC 60  
 1 ATCTTGACAGATTAAATCGAATCTTCTTCACTGAGCGGTACAGTATCGAGAGGGTAATC 60

61 ATGCGGTGCTGATTCGACGAGACCAAAACAAAGAAATTCGGGGTGCATCACCCG 120  
 61 ATGCGGTGCTGATTCGACGAGACCAAAACAAAGAAATTCGGGGTGCATCACCCG 120

61 ATGCGGTGCTGATTCGACGAGACCAAAACAAAGAAATTCGGGGTGCATCACCCG 120  
 61 ATGCGGTGCTGATTCGACGAGACCAAAACAAAGAAATTCGGGGTGCATCACCCG 120

121 GCGGGGTGCGGAGACTAACCGGTGCGCATGAGTCTCATCGAGCAGATGCGCGA 180  
 121 GCGGGGTGCGGAGACTAACCGGTGCGCATGAGTCTCATCGAGCAGATGCGCGA 180

121 GCGGGGTGCGGAGACTAACCGGTGCGCATGAGTCTCATCGAGCAGATGCGCGA 180  
 121 GCGGGGTGCGGAGACTAACCGGTGCGCATGAGTCTCATCGAGCAGATGCGCGA 180

181 GAGGGCTGCGCTATACACGACGCGGATTTCAAGGCGGACGCGCACTGCTGCGCAC 240  
 181 GAGGGCTGCGCTATACACGACGCGGATTTCAAGGCGGACGCGCACTGCTGCGCAC 240

181 GAGGGCTGCGCTATACACGACGCGGATTTCAAGGCGGACGCGCACTGCTGCGCAC 240  
 181 GAGGGCTGCGCTATACACGACGCGGATTTCAAGGCGGACGCGCACTGCTGCGCAC 240

241 GCGGACGAGGTGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGAACGATAGCGCG 300  
 241 GCGGACGAGGTGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGAACGATAGCGCG 300

241 GCGGACGAGGTGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGAACGATAGCGCG 300  
 241 GCGGACGAGGTGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGAACGATAGCGCG 300

301 GAATACGCGCGCTGCGACGAGCGGATCTGCTTCACTTCTGATTTGCGCGGCTCAC 360  
 301 GAATACGCGCGCTGCGACGAGCGGATCTGCTTCACTTCTGATTTGCGCGGCTCAC 360

301 GAATACGCGCGCTGCGACGAGCGGATCTGCTTCACTTCTGATTTGCGCGGCTCAC 360  
 301 GAATACGCGCGCTGCGACGAGCGGATCTGCTTCACTTCTGATTTGCGCGGCTCAC 360

361 GTGCTTGACAGCGATGCGTGTGATTCGCGGACACGATCAATGGCTACGAGACCGTCC 420  
 361 GTGCTTGACAGCGATGCGTGTGATTCGCGGACACGATCAATGGCTACGAGACCGTCC 420

361 GTGCTTGACAGCGATGCGTGTGATTCGCGGACACGATCAATGGCTACGAGACCGTCC 420  
 361 GTGCTTGACAGCGATGCGTGTGATTCGCGGACACGATCAATGGCTACGAGACCGTCC 420

421 AAGCGCGGAGGCGGACGACGCGGCTGCGCGGATGCGGAGGCGGAGGCGGCGACTG 480  
 421 AAGCGCGGAGGCGGACGACGCGGCTGCGCGGATGCGGAGGCGGAGGCGGCGACTG 480

421 AAGCGCGGAGGCGGACGACGCGGCTGCGCGGATGCGGAGGCGGAGGCGGCGACTG 480  
 421 AAGCGCGGAGGCGGACGACGCGGCTGCGCGGATGCGGAGGCGGAGGCGGCGACTG 480

481 CCGCGGAGGTGGCGCTTACCACTGATGCGAACCCAAAGGGGCGCGGCTGCTGATG 540  
 481 CCGCGGAGGTGGCGCTTACCACTGATGCGAACCCAAAGGGGCGCGGCTGCTGATG 540

481 CCGCGGAGGTGGCGCTTACCACTGATGCGAACCCAAAGGGGCGCGGCTGCTGATG 540  
 481 CCGCGGAGGTGGCGCTTACCACTGATGCGAACCCAAAGGGGCGCGGCTGCTGATG 540

541 GCGGGGTGCGGCGGCTGCGAACGCGGACGATGCTGATTCGCGCGGCGGACGCGCGCT 600  
 541 GCGGGGTGCGGCGGCTGCGAACGCGGACGATGCTGATTCGCGCGGCGGACGCGCGCT 600

541 GCGGGGTGCGGCGGCTGCGAACGCGGACGATGCTGATTCGCGCGGCGGACGCGCGCT 600  
 541 GCGGGGTGCGGCGGCTGCGAACGCGGACGATGCTGATTCGCGCGGCGGACGCGCGCT 600

601 ACAAGCGAGCGGCGATGCGCAAGCGGATGCGGCGGCGGACGCTTACGATAGACATCA 660  
 601 ACAAGCGAGCGGCGATGCGCAAGCGGATGCGGCGGCGGACGCTTACGATAGACATCA 660

601 ACAAGCGAGCGGCGATGCGCAAGCGGATGCGGCGGCGGACGCTTACGATAGACATCA 660  
 601 ACAAGCGAGCGGCGATGCGCAAGCGGATGCGGCGGCGGACGCTTACGATAGACATCA 660

661 TCGACAAACTTGGGCAACTGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 720  
 661 TCGACAAACTTGGGCAACTGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 720

661 TCGACAAACTTGGGCAACTGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 720  
 661 TCGACAAACTTGGGCAACTGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 720

721 CCGGCTTACGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 780  
 721 CCGGCTTACGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 780

721 CCGGCTTACGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 780  
 721 CCGGCTTACGAGCTGCGAGGCTGCGCAAGCGGATGCGGCGGCGGATGCACTGCTGATCAT 780

781 TCGCAGGCGCGCAAGGCGCAAAATAGTCTGCAATCTGCTGCGGATGCAATGCAACAG 840  
 781 TCGCAGGCGCGCAAGGCGCAAAATAGTCTGCAATCTGCTGCGGATGCAATGCAACAG 840

781 TCGCAGGCGCGCAAGGCGCAAAATAGTCTGCAATCTGCTGCGGATGCAATGCAACAG 840  
 781 TCGCAGGCGCGCAAGGCGCAAAATAGTCTGCAATCTGCTGCGGATGCAATGCAACAG 840

841 GTGCGGTGCTGATGATAGCCATGACGAGGCGGCTGTTGCAAGGCTCACAGCGCA 900  
 841 GTGCGGTGCTGATGATAGCCATGACGAGGCGGCTGTTGCAAGGCTCACAGCGCA 900

841 GTGCGGTGCTGATGATAGCCATGACGAGGCGGCTGTTGCAAGGCTCACAGCGCA 900  
 841 GTGCGGTGCTGATGATAGCCATGACGAGGCGGCTGTTGCAAGGCTCACAGCGCA 900

901 CCGCCTTACGAGCGGCGGATGCGGCGGATGCGGCGGCGGATGCACTGCTGATCAT 960  
 901 CCGCCTTACGAGCGGCGGATGCGGCGGATGCGGCGGCGGATGCACTGCTGATCAT 960

901 CCGCCTTACGAGCGGCGGATGCGGCGGATGCGGCGGCGGATGCACTGCTGATCAT 960  
 901 CCGCCTTACGAGCGGCGGATGCGGCGGATGCGGCGGCGGATGCACTGCTGATCAT 960

961 TCGCAGGCGCGCAAGGCGCAAGGCGGATGCGGCGGATGCGGCGGATGCACTGCTGATCAT 1020  
 961 TCGCAGGCGCGCAAGGCGCAAGGCGGATGCGGCGGATGCGGCGGATGCACTGCTGATCAT 1020

961 TCGCAGGCGCGCAAGGCGCAAGGCGGATGCGGCGGATGCGGCGGATGCACTGCTGATCAT 1020  
 961 TCGCAGGCGCGCAAGGCGCAAGGCGGATGCGGCGGATGCGGCGGATGCACTGCTGATCAT 1020

1021 TCGCTGAGCTTGGCGACATGCTGCGGCGGCGGCTGCTGCTGCAATCCGCGCACTAGCCA 1080  
 1021 TCGCTGAGCTTGGCGACATGCTGCGGCGGCGGCTGCTGCTGCAATCCGCGCACTAGCCA 1080

```

Db      421 CAGACCCGCGACGCGACACTACCCCTGCTTGCCCGATGAGCGAAGTCCGGGTGCACTC 480
QY      480 GCGGCCAGTTGGGCGCTTACCACTGATGCCGAACCCAGGGGGCGCGGTGCTGATG 539
Db      481 GCGGCCAGTTGGGCGCTTACCACTGATGCCGAACCCAGGGGGCGCGGTGCTGATG 540
QY      540 GCGGCCAGTTGGGCGCTTACCACTGATGCCGAACCCAGGGGGCGCGGTGCTGATG 599
Db      541 GCGGCCAGTTGGGCGCTTACCACTGATGCCGAACCCAGGGGGCGCGGTGCTGATG 600
QY      600 TACACGACGCCGCAATCCCAACGGCATGCGCGCAACGCTTACGCTTACGACATCAAC 659
Db      601 TACACGACGCCGCAATCCCAACGGCATGCGCGCAACGCTTACGCTTACGACATCAAC 660
QY      660 ATGACAACTTCGAGACTCGACGCGGATTCGCGCGCGGATCCACATCCTCTACTA 719
Db      661 ATGACAACTTCGAGACTCGACGCGGATTCGCGCGCGGATCCACATCCTCTACTA 720
QY      720 TCGGCTTACGAGCTGAGGGTGCCTCAACAGTGCAGACCTGGTATGCGGCGCTCTG 779
Db      721 TCGGCTTACGAGCTGAGGGTGCCTCAACAGTGCAGACCTGGTATGCGGCGCTCTG 780
QY      780 GTGCGAGGCGCAAGGACCCCAATAGTCTGCAATTCCTGCTGCGGCAATGAAACCA 839
Db      781 GTGCGAGGCGCAAGGACCCCAATAGTCTGCAATTCCTGCTGCGGCAATGAAACCA 840
QY      840 GGTGCGGCTAGTGTGATATAGCCATCGACGAGCGGCGCTGTTTCGAAGGCTCAGCAGCG 899
Db      841 GGTGCGGCTAGTGTGATATAGCCATCGACGAGCGGCGCTGTTTCGAAGGCTCAGCAGCG 900
QY      900 ACCACTAGACCAACCGGAGCTTCCGCTGCGACAGACAGCTGTTTACTGCTGCGGCAAC 959
Db      901 ACCACTAGACCAACCGGAGCTTCCGCTGCGACAGACAGCTGTTTACTGCTGCGGCAAC 960
QY      960 ATGCGCGCTCGGTGCGCAAGAGCTGACCTGACGCTGACCAACGCGAGATGCCGTAT 1019
Db      961 ATGCGCGCTCGGTGCGCAAGAGCTGACCTGACGCTGACCAACGCGAGATGCCGTAT 1020
QY      1020 GTGCTGAGCTTCCGACCAATGCTGCGGCGCGCTGCGGCTGCAATCCGACATAGCC 1079
Db      1021 GTGCTGAGCTTCCGACCAATGCTGCGGCGCGCTGCGGCTGCAATCCGACATAGCC 1080
QY      1080 AAGGCTTTCGACGCGACGAGAGGGGGGCTTACTGTCGAGAGGGTGGCCACCGACCTGCGG 1139
Db      1081 AAGGCTTTCGACGCGACGAGAGGGGGGCTTACTGTCGAGAGGGTGGCCACCGACCTGCGG 1140
QY      1140 GTGCGCTTACCGAGCGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1199
Db      1141 GTGCGCTTACCGAGCGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
QY      1200 CACACTCGGAGTAGGAGGAGGATGATGCTGCGCC 1235
Db      1201 CACACTCGGAGTAGGAGGAGGATGATGCTGCGCC 1236

```

RESULT 8  
 LOCUS A89753 1236 bp DNA  
 DEFINITION A89753 Sequence 10 from Patent WO9832862. PAT 22-JAN-2000  
 VERSION A89753.1 GI:6738287  
 KEYWORDS  
 ORGANISM  
 SOURCE  
 REFERENCES  
 1 (bases 1 to 1236)  
 Flohe, L. and Singh, M.  
 L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
 Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE);  
 FEATURES  
 source 1..1236  
 /organism="unidentified"

```

BASE COUNT      236 a      395 c      385 g      220 t
ORIGIN
Query Match      99.0%; Score 1222.4; DB 5; Length 1236;
Best local similarity 99.8%; Pred. 1.7e-166;
Matches 1234; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```

/db\_xref="taxon:32644"

```

QY      1 ATCTTCAGATTATATCAACTTCTTACACATGAAACGATACAGTATCGAAGGGTATATC 60
Db      1 ATCTTCAGATTATATCAACTTCTTACACATGAAACGATACAGTATCGAAGGGTATATC 60
QY      61 ATGCGCGTGGTATTCGCGCGGAGCAAAACCAAGATTCGGGGTGCATACCCCG 120
Db      61 ATGCGCGTGGTATTCGCGCGGAGCAAAACCAAGATTCGGGGTGCATACCCCG 120
QY      121 GCGGCGTGCAGCAATCAACCGTGTGCGCATAGGTGCTATCCAGGAGGAGCGCGA 180
Db      121 GCGGCGTGCAGCAATCAACCGTGTGCGCATAGGTGCTATCCAGGAGGAGCGCGA 180
QY      181 GAGGCTCGGCTATACCGACGCGGATTTCAAGCGCGAGCGCGCAACTGCTGCGCAC 240
Db      181 GAGGCTCGGCTATACCGACGCGGATTTCAAGCGCGAGCGCGCAACTGCTGCGCAC 240
QY      241 GCGACACAGGTGTGGCGCGAGCTGATTTATGTCMAAGTCAAGAAACCGATACCGGG 300
Db      241 GCGACACAGGTGTGGCGCGAGCTGATTTATGTCMAAGTCAAGAAACCGATACCGGG 300
QY      301 GAAT..CGGCGGCTCGACACAGGCG..GATCTTGTTCAGCTTCTGCAATTGCGCGCTCA 359
Db      301 GAAT..CGGCGGCTCGACACAGGCG..GATCTTGTTCAGCTTCTGCAATTGCGCGCTCA 360
QY      360 CGTGTTCAGCGATGCGTGTGTTGATTCGCGACACGCTCAATTGCTTACGACACCGTC 419
Db      361 CGTGTTCAGCGATGCGTGTGTTGATTCGCGACACGCTCAATTGCTTACGACACCGTC 420
QY      420 CAGACCGCGAGAGGCGCACTACCTGCTTGCCTGCGGATAGGAGGATGCGCGGTGAGCTC 479
Db      421 CAGACCGCGAGAGGCGCACTACCTGCTTGCCTGCGGATAGGAGGATGCGCGGTGAGCTC 480
QY      480 GCGCGCCAGGTTGCGCTTACCACTGATCGAACCAGAGGGGGCGCGGTGCTGATG 539
Db      481 GCGCGCCAGGTTGCGCTTACCACTGATCGAACCAGAGGGGGCGCGGTGCTGATG 540
QY      540 GCGGGGTGCGCGGCGTCAACCGGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 599
Db      541 GCGGGGTGCGCGGCGTCAACCGGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
QY      600 TACACGACGCCGCGATCGCAACGGCATGCGCGGACGCTTACGCTTACGATCAAC 659
Db      601 TACACGACGCCGCGATCGCAACGGCATGCGCGGACGCTTACGCTTACGATCAAC 660
QY      660 ATGACAACTTCGCGCACTGACGCGGATTCGCGCGCGGATCCACATCCTCTACTA 719
Db      661 ATGACAACTTCGCGCACTGACGCGGATTCGCGCGCGGATCCACATCCTCTACTA 720
QY      720 TCGGCTTACGAGCTGAGGGTGCCTCAACAGTGCAGACCTGGTATGCGGCGCTCTG 780
Db      721 TCGGCTTACGAGCTGAGGGTGCCTCAACAGTGCAGACCTGGTATGCGGCGCTCTG 780
QY      780 GTGCGAGGCGCAAGGACCCCAATAGTCTGCAATTCCTGCTGCGGCAATGAAACCA 839
Db      781 GTGCGAGGCGCAAGGACCCCAATAGTCTGCAATTCCTGCTGCGGCAATGAAACCA 840
QY      840 GGTGCGGCTAGTGTGATATAGCCATCGACGAGCGGCTGTTTCGAAGGCTCAGCAGCG 899
Db      841 GGTGCGGCTAGTGTGATATAGCCATCGACGAGCGGCTGTTTCGAAGGCTCAGCAGCG 900
QY      900 ACCACTAGACCAACCGGAGCTTCCGCTGCGACAGACAGCTGTTTACTGCTGCGGCAAC 959
Db      901 ACCACTAGACCAACCGGAGCTTCCGCTGCGACAGACAGCTGTTTACTGCTGCGGCAAC 960

```



QY 960 ATGCCCGCTCGGTGCGCAAGACGTGACCTGACGCGCTGACCAACGCGAGTATCCGCTAT 1019  
 |||||  
 Db 961 ATGCCCGCTCGGTGCGCAAGACGTGACCTGACGCGCTGACCAACGCGAGTATCCGCTAT 1020  
 |||||  
 QY 1020 GTGCTGACGCTTCCGACCATGCTGCGCGCGCGCTGCGGTGAAATCCGCGACTAGCC 1079  
 |||||  
 Db 1021 GTGCTGACGCTTCCGACCATGCTGCGCGCGCGCTGCGGTGAAATCCGCGACTAGCC 1080  
 |||||  
 QY 1080 AAAGGTCTTTCAGACGCAAGGAGGCGTTACTGTCCGAAGGGGTGGCCACGACCTGGGG 1139  
 |||||  
 Db 1081 AAAGGTCTTTCAGACGCAAGGAGGCGTTACTGTCCGAAGGGGTGGCCACGACCTGGGG 1140  
 |||||  
 QY 1140 GTGCGCTTCACGACGACCGCGCGCGCGCTGCTGCTGACTCTCGGCGCGCTGTTACGCGGAG 1199  
 |||||  
 Db 1141 GTGCGCTTCACGACGACCGCGCGCGCGCTGCTGCTGACTCTCGGCGCGCTGTTACGCGGAG 1200  
 |||||  
 QY 1200 CACAGCTGGGAGTAAGGAAGGAGGATGATGCGGCC 1235  
 |||||  
 Db 1201 CACAGCTGGGAGTAAGGAAGGAGGATGATGCGGCC 1236  
 |||||

RESULT 9  
 LOCUS A87607 1237 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 4 from Patent W09836089.  
 ACCESSION A87607  
 VERSION A87607.1 GI:6736247  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1237)  
 FIOHE, L. and Singh, M.  
 TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
 FIOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source 1..1237  
 location/Qualifiers  
 1..1237  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 236 a 394 c 386 g 221 t  
 ORIGIN

Query Match 98.9%; Score 1220.8; DB 5; Length 1237;  
 Best Local Similarity 99.8%; Pred. No. 2.9e-166;  
 Matches 1233; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 ATCTTGAGATTATGCACTTCTTACAGTGAAGGCTACAGTATGAGAGGGTATTC 60  
 |||||  
 Db 1 ATCTTGAGATTATGCACTTCTTACAGTGAAGGCTACAGTATGAGAGGGTATTC 60  
 |||||  
 QY 61 ATGCGGCTCGGTATTCGACGACGACCAAAACAAAGAAATCCGGGTGCGCATACCCCG 120  
 |||||  
 Db 61 ATGCGGCTCGGTATTCGACGACGACCAAAACAAAGAAATCCGGGTGCGCATACCCCG 120  
 |||||  
 QY 121 GCCGCGCTCGGAACTAACCCGTCGTGCGCATGAGGTGCTCATCCAGGAGGTGCGGGA 180  
 |||||  
 Db 121 GCCGCGCTCGGAACTAACCCGTCGTGCGCATGAGGTGCTCATCCAGGAGGTGCGGGA 180  
 |||||  
 QY 181 GAGGGCTCGGCTATCAGCAGCGCGGATTTCAAGGGGCGAGCGCGCACTGTGCGGACC 240  
 |||||  
 Db 181 GAGGGCTCGGCTATCAGCAGCGCGGATTTCAAGGGGCGAGCGCGCACTGTGCGGACC 240  
 |||||  
 QY 241 GCGGACCGAGTGTGGGCGCGCGCTGATTTATGCTCAAGGTCAAAAGACCGATAGCGGGG 300  
 |||||  
 Db 241 GCGGACCGAGTGTGGGCGCGCGCTGATTTATGCTCAAGGTCAAAAGACCGATAGCGGGG 300  
 |||||  
 QY 301 GAATACGCGCGCTGCGACACGCGGCG -GATCTTGTCAAGTCTTGTGATTTGGCGCGCTCA 359  
 |||||  
 Db 301 GAATACGCGCGCTGCGACACGCGGCGAGATCTTGTGATTTGTGATTTGGCGCGCTCA 360  
 |||||  
 QY 360 CGTGCTGACCAAGATGCGTGTGATTCGCGGACCAAGCAATGCTTACGAGACCGCTC 419  
 |||||

Db 361 CGTGCTGACCAAGATGCGTGTGATTCGCGGACCAAGCAATGCTTACGAGACCGCTC 420  
 |||||  
 QY 420 CAGACCGCGGAAAGGCGCACTACCTGCTTGGCCCGATGAGCGAAGTCCGCGTGAATC 479  
 |||||  
 Db 421 CAGACCGCGGAAAGGCGCACTACCTGCTTGGCCCGATGAGCGAAGTCCGCGTGAATC 480  
 |||||  
 QY 480 GCGCGCGCGGCTTGTGCGCTTACCACTGATGCGAACCAGGAGGCGCGGTGTCTGATG 539  
 |||||  
 Db 481 GCGCGCGCGGCTTGTGCGCTTACCACTGATGCGAACCAGGAGGCGCGGTGTCTGATG 540  
 |||||  
 QY 540 GCGCGGCTGCGCGCGCGCGCAAGCGCGCGCGCTGCTGATGCGCGCGCGCGCGCGCG 599  
 |||||  
 Db 541 GCGCGGCTGCGCGCGCGCGCAAGCGCGCGCGCTGCTGATGCGCGCGCGCGCGCGCG 600  
 |||||  
 QY 600 TACAAACGACCGCGCATGCGCAACGCGCATGGCGCGGACCGTTAGGTTCTAGACATCAAC 659  
 |||||  
 Db 601 TACAAACGACCGCGCATGCGCAACGCGCATGGCGCGGACCGTTAGGTTCTAGACATCAAC 660  
 |||||  
 QY 660 ATCGACAAACTTGGGCACTGACGCGCGAGTCTGCGCGCGGATGCACTGCGTACTCA 719  
 |||||  
 Db 661 ATCGACAAACTTGGGCACTGACGCGCGAGTCTGCGCGCGGATGCACTGCGTACTCA 720  
 |||||  
 QY 720 TCGGCTTACGACCTGAGAGGTGCGCTCAACGCTGCGCATGCTGATTTGGGCGGCTG 779  
 |||||  
 Db 721 TCGGCTTACGACCTGAGAGGTGCGCTCAACGCTGCGCATGCTGATTTGGGCGGCTG 780  
 |||||  
 QY 780 GTGCCAGGCGCCCAAGGCGACCCCAATTAATCTGCAATTAATCTGCGCGCATATGAACCA 839  
 |||||  
 Db 781 GTGCCAGGCGCCCAAGGCGACCCCAATTAATCTGCAATTAATCTGCGCGCATATGAACCA 840  
 |||||  
 QY 840 GGTGCGTACTGCTGATATAGCCATGACGACGAGGCGGCTTTGCAAGGCTCAGCAGCG 899  
 |||||  
 Db 841 GGTGCGTACTGCTGATATAGCCATGACGACGAGGCGGCTTTGCAAGGCTCAGCAGCG 900  
 |||||  
 QY 900 ACCACTTACGACCGACCGACCGCTGCGCGCTGACGACGACGCTTTTACTGCTGCGCAAC 959  
 |||||  
 Db 901 ACCACTTACGACCGACCGACCGCTGCGCGCTGACGACGACGCTTTTACTGCTGCGCAAC 960  
 |||||  
 QY 960 ATGCCCGCTCGGTGCGCAAGACGTGACCTACGCGCTGACGACGACGAGCATGCGCTAT 1019  
 |||||  
 Db 961 ATGCCCGCTCGGTGCGCAAGACGTGACCTACGCGCTGACGACGACGAGCATGCGCTAT 1020  
 |||||  
 QY 1020 GTGCTGACGCTTGGCGACCATGCTGCGCGCGCGCTGCGGATCCGCGACTAGCC 1079  
 |||||  
 Db 1021 GTGCTGACGCTTGGCGACCATGCTGCGCGCGCGCTGCGGATCCGCGACTAGCC 1080  
 |||||  
 QY 1080 AAAGGTCTTTCAGACGCAAGGAGGCGCTTACTGTCCGAAGGGGTGGCCACGACCTGGGG 1139  
 |||||  
 Db 1081 AAAGGTCTTTCAGACGCAAGGAGGCGCTTACTGTCCGAAGGGGTGGCCACGACCTGGGG 1140  
 |||||  
 QY 1140 GTGCGCTTACCGACGCGCGCGCGCGCTGCTGCTGACTCTGCGCGGCTGTTACGCGGAG 1199  
 |||||  
 Db 1141 GTGCGCTTACCGACGCGCGCGCGCGCTGCTGCTGACTCTGCGCGGCTGTTACGCGGAG 1200  
 |||||  
 QY 1200 CACAGCTGGGAGTAAGGAAGGAGGATGATGCGGCC 1235  
 |||||  
 Db 1201 CACAGCTGGGAGTAAGGAAGGAGGATGATGCGGCC 1236  
 |||||

RESULT 10  
 LOCUS A89747 1237 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 4 from Patent W09832862.  
 ACCESSION A89747  
 VERSION A89747.1 GI:6738281  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1237)  
 FIOHE, L. and Singh, M.  
 TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM



JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES Location/Qualifiers  
 Source 1.1237  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 236 a 394 c 386 g 221 t  
 ORIGIN

Query Match 98.9%; Score 1220.8; DB 5; Length 1237;  
 Best Local Similarity 99.8%; Pred. No. 2.9e-166;  
 Matches 1233; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```

QY 1 ATCTTGAGATTAATGAACTTTCTTACACATGAAGGTACAGTATGAGAGGGTAAATC 60
DB 1 ATCTTGAGATTAATGAACTTTCTTACACATGAAGGTACAGTATGAGAGGGTAAATC 60
QY 61 ATGCGGCTGCTATTCGACGACGACCAAAACAAAGAAATCCGGGTGGCATCACCCG 120
DB 61 ATGCGGCTGCTATTCGACGACGACCAAAACAAAGAAATCCGGGTGGCATCACCCG 120
QY 121 GCCGCGCTCGGAGAACTAACCCCTGCTGAGCATGAGTGTCTCATCCAGCAGGTGCGGA 180
DB 121 GCCGCGCTCGGAGAACTAACCCCTGCTGAGCATGAGTGTCTCATCCAGCAGGTGCGGA 180
QY 181 GAGGCTCGCTATTCACGACGCGGATTCAGGGCGGAGCGCGCACTGTGCGCAC 240
DB 181 GAGGCTCGCTATTCACGACGCGGATTCAGGGCGGAGCGCGCACTGTGCGCAC 240
QY 241 GCGGACGAGTGTGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGACGATAGGGGG 300
DB 241 GCGGACGAGTGTGGGCGGACGCTGATTTATGCTCAAGGTCAAAAGACGATAGGGGG 300
QY 301 GAATACGCGCCCTGCGACACAGGGG-GATCTTGTTCACGTTTGTGATTTGGCGCGTCA 359
DB 301 GAATACGCGCCCTGCGACACAGGGG-GATCTTGTTCACGTTTGTGATTTGGCGCGTCA 359
QY 360 CGTGTGACGACGATCGTGTGTGATTCGCGGACACGTCATTTGCCATACGACCGTTC 419
DB 360 CGTGTGACGACGATCGTGTGTGATTCGCGGACACGTCATTTGCCATACGACCGTTC 419
QY 420 CAGACGCGGAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 479
DB 420 CAGACGCGGAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 479
QY 480 GCGGCGGAGTGTGGGCGCTTACACGCTGATGCGAAAGGAGGCGCGGTGTGTGATG 539
DB 480 GCGGCGGAGTGTGGGCGCTTACACGCTGATGCGAAAGGAGGCGCGGTGTGTGATG 539
QY 540 GGGGGGCTCGCGGCTGCAACCGGCGGACGCTGATGCGGCGCGGCGGCGGCGGCGG 599
DB 540 GGGGGGCTCGCGGCTGCAACCGGCGGACGCTGATGCGGCGCGGCGGCGGCGGCGGCGG 599
QY 600 TACAAGCGACCGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 659
DB 600 TACAAGCGACCGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 659
QY 660 ATGCAAACTTGGCACTGACGCGGAGTTGTGGGCG GATCAGCACTGCTCTA 719
DB 660 ATGCAAACTTGGCACTGACGCGGAGTTGTGGGCG GATCAGCACTGCTCTA 719
QY 720 TCGGCTTACGAGTGTGAGGAGTGTGCGCAAGTGTGCTGATTTGGGCGGCTG 779
DB 720 TCGGCTTACGAGTGTGAGGAGTGTGCGCAAGTGTGCTGATTTGGGCGGCTG 779
QY 780 GTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 839
DB 780 GTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 839
QY 840 GGTGCGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 899
DB 840 GGTGCGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 899

```

```

QY 900 ACCACCTACGACACCGGAGCTTGGCGGCGGACGACGACGCTTTACTGCTGGCGAC 959
DB 901 ACCACCTACGACACCGGAGCTTGGCGGCGGACGACGACGCTTTACTGCTGGCGAC 960
QY 960 ATGCGGCGCTCGGTCGCGGAGGAGCTGACGACGCTGACGACGACGACGACGCTAT 1019
DB 961 ATGCGGCGCTCGGTCGCGGAGGAGCTGACGACGCTGACGACGACGACGACGCTAT 1020
QY 1020 GTGCTGACGCTTGGGACGACGATGCTGCGGCGGCGGCTGCGGCTGCAATCGGCTAGCC 1079
DB 1021 GTGCTGACGCTTGGGACGACGATGCTGCGGCGGCGGCTGCGGCTGCAATCGGCTAGCC 1080
QY 1080 AAAGGCTTTCGACGACGACGAGGCGGCTGCTGCTGCGGAGGCGGCGGCGGCGGCGG 1139
DB 1081 AAAGGCTTTCGACGACGACGAGGCGGCTGCTGCTGCGGAGGCGGCGGCGGCGGCGG 1140
QY 1140 GTGCGGCTTACGACGACGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1199
DB 1141 GTGCGGCTTACGACGACGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
QY 1200 CACAGCTGCGGAGTAAAGGAGGAGGATGATGCTGGCC 1235
DB 1201 CACAGCTGCGGAGTAAAGGAGGAGGATGATGCTGGCC 1236

```

## RESULT 11

MTV002 56414 bp DNA BCT 17-JUN-1998  
 LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.  
 ACCESSION AL008967 AL123435  
 VERSION AL008967.1 GI:3261491  
 KEYWORDS

## SOURCE

ORGANISM Mycobacterium tuberculosis.  
 Bacteria; Firmi; tes; Actinobacteriia; Actinobacteriidae;  
 Actinomycetales; Corynebacteriineae; Mycobacteriaceae;

## REFERENCE

AUTHORS Cole, S.T., Broesch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E., Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., O'Brien, A.D., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S., Squares, S., Squires, R., Sultson, J.E., Taylor, K., Whitehead, S. and Barrall, B.G.  
 TITLE Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence  
 JOURNAL Nature 393 (6685), 537-544 (1998)  
 MEDLINE 98295987  
 REMARK Erratum: [[published erratum appears in Nature 1998 Nov 12;396(6707):190]]  
 2 (bases 1 to 56414)

## REFERENCE

AUTHORS Parkhill, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
 COMMENT On Jun 27, 1998 this sequence version replaced GI:3261491.  
 Note(s): Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.  
 (URL, <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.  
 Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.

CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

## FEATURES

```

source
    Location/Qualifiers
        1..56414
        /organism="Mycobacterium tuberculosis"
        /strain="H37Rv"
        /db_xref="taxon:1773"
        <1..>56414
        /note="fragment designated v002. Does not represent a physical clone"
        <1..233
        /organism="Mycobacterium tuberculosis"
        /strain="H37Rv"
        /db_xref="taxon:1773"
        /clone="Y154"
        complement(3..527)
        /gene="recX"
        complement(3..527)
        /gene="recX"
        /note="Rv2736c, (MTV002.01c), len: 174 aa. recX, similar to eg. RECX PSEAE P37860 regulatory protein recX from Pseudomonas aeruginosa (153 aa), fasta scores: opt: 161 z-score: 257.2 E(): 3.6e-07, 30.7% identity in 137 aa overlap. Overlaps and extends CDS from overlapping cosmid MTCY154.16c"
        /codon_start=1
        /transl_table=11
        /product="recX"
        /protein_id="CAI5532.1"
        /db_xref="GI:2624257"
        /db_xref="SPTREMBL:O33280"
        /translation="MTVSCPPSTSEREEQARALCLLTARSTRRAELAGOLAKRGY PEDIGNVLDRLAAGLVDDTDFAEQVQSRANAAKSRALAEHLAKVDDVITV VLGGIDGAERGAERKLVARARLREVLIDGTDEARVSRRLVAMLARVGGQLACEV VIELAERERRRRV"
        complement(493..2865)
        /gene="recA"
        complement(493..2865)
        /gene="recA"
        /note="Rv2737c, (MTV002.02c), len: 790 aa. recA, identical to RECA_MYTU P26345 recA protein (790 aa). Contains self-splicing protein element (intron) from 2294 to 969 (c), similar to intron II from PR:E32317 (EMBL:Y13030) DNA-directed DNA polymerase (PC 2.7.7.7) from Thermococcus sp. (1829 aa), fasta scores: opt: 81 z-score: 235.2 E(): 6e-06, 24.6% identity in 183 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop), PS00321 recA signature, and PS00881 protein splicing signature. See Davis et al, (1992) Cell 71(2):201-210"
        /codon_start=1
        /transl_table=11
        /product="recA"
        /protein_id="CAI5533.1"
        /db_xref="GI:2624258"
        /db_xref="SWISS-PROT:P26345"
        /translation="MTQTPDREKALAAVAOIKSYGKGSVMRLGDEARPIVITG SIALDVALIGTGLPGRVTEIYGPSSGKTTVALHVAANAQAAGVAALFIDAHADP DYAKKLGVDTSLLVSOPDTGEQALEADLISGALDIVIDVSALVPRALLEGEM GDSHVGLQARLSQALRKMTGALNNGSTAIFINLRDKIGVMSPEPTTGKAKLF YASVRADVREVELTKDGINAVGNRTKVKVKNKCLAEGRIFEDPTGTHRIEDVDG RKPIHVAANKDCTILHAPVSWFDGTRVIGLRAGGAIWATPDHVLVEYDVGRA ALGRKGDRAVQPRRFQSDSAPADAHARILGILIGSGRWGCKPIMPINOR ALIDDTRIATAGCAHPGRSISIAHRRGCRNGVALDCQAGYKGLAKMTKIPN WREPELADIYGNLIFGLFESDGMVSRDTGLRVGYTTTSQLAHQJHMLLRFV GSTVYRDPQKRPSTVNGRRIDSKHOFVEVRISGNDVTAFAESVPMGPPRAALIQ ALPEATQGRRRSQATYLAEMTDVALENTLDERGVTAQEAAMIGVAGSDPRGMKOV LGASRLRDRVQALADLDKFLHMLAEELRSYREVLPTRARPTFLEVEELHTL VAGVYVHNSPPFKOAEFDIILYKGISRSRGLSDMGVOGLTKRSKGFTEGSEGLS OGKRNARNFLVENADVADIEKKIKKELGIGAVYTDPSNDGVLPAVDF"

```

```

misc_feature
    complement(790..2112)
    /gene="recA"
    /note="recA intron"
    complement(790..807)
    /gene="recA"
    /note="PS00881 protein splicing signature"
    complement(2197..2223)
    /gene="recA"
    /note="PS00321 recA signature"
    complement(2644..2667)
    /gene="recA"
    /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
    complement(2873..2877)
    /note="possible RBS upstream of Rv2737c"
    complement(3247..3453)
    /gene="Rv2738c"
    complement(3247..3453)
    /gene="Rv2738c"
    /note="Rv2738c, (MTV002.03c), len: 68 aa. Unknown, N-terminus is highly similar to the N-terminus of the upstream ORF MTV002.0/c (78.4% identity in 37 aa overlap); also similar to A020958|SC4H_5 Streptomyces coelicolor cosmid 4H8 (64aa) opt: 185 z-score: 283.5 E(): 2.9e-08; 39.7% identity in 63 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein Rv2738c"
    /protein_id="CAI5534.1"
    /db_xref="GI:2624260"
    /db_xref="SPTREMBL:O33281"
    /translation="MLAGVRLTEFHERVALHFGAAYGSSVLLDHYLTGDFGRSAQAQI EDGVEPRDYWRALCADFDVPHDRW"
    complement(3464..4630)
    /gene="Rv2739c"
    complement(3464..4630)
    /gene="Rv2739c"
    /note="Rv2739c, (MTV002.04c), len: 388 aa. Probable trans. use, similar to eg. TR:O51560 (EMBL:L28170) Pseudom. as aeruginosa thiamosyl transferase (442 aa), fasta scores: opt: 178 z-score: 226.3 E(): 1.9e-05, 25.9% identity in 425 aa overlap. Equivalent to Mycobacterium leprae protein MLCB33.02c (392 aa), fasta scores: 9P1294723|MLCB33.2 Mycobacterium leprae cosmid B33 opt: 2112 z-score: 2364.5 E(): 0; 80.9% identity in 388 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein Rv2739c"
    /protein_id="CAI5535.1"
    /db_xref="GI:2624261"
    /db_xref="SPTREMBL:O33282"
    /translation="WRVAVVAGPDPGHSPFALICQRRRAADPTLTFTGVEMEAR AAGIDAVELDGLAATRDLDAGARLHRRAAQAVALNVRRLALELTVSDVITACGG MAALIGIPWELNHPRLPLPSKGLPISGIAVGTJRGRLDRLATLGRSBRG LRORAARVREIIGLPARDGPLRLIATLPALEVPDPDPAVAVVGPILHFPPTRYLA IPAGGPVAVVAPSTALGTAGTLEVALOSTTPTGTVSGSRVYVSRISGADLVYPPH AVAGISQAELTRADLVITCGGHRKAKTLLAGVPVYVGGGQWETIARVYRQGS AVLIRPLTADLAAVAVNEVLSSPRREARRRAASVAGADPVRCBALALAG"
    4674..5123
    /gene="Rv2740"
    4674..5123
    /gene="Rv2740"
    /note="Rv2740, (MTV002.05), len: 149 aa. Unknown, similar to M. leprae hypothetical protein TR:049950 (EMBL:294723) MLCB33.03 (82253-53,140) (178 aa), fasta scores: opt: 498 z-score: 538.9 E(): 7.3e-23, 51.6% identity in 161 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein Rv2740"
    /protein_id="CAI5536.1"
    /db_xref="GI:2624262"
    /db_xref="SPTREMBL:O33283"

```

/translation="MAELTETSPETPETTEIRAIVEAFLNALQNEDEFTDALGDL  
VYENVGSRIRGRRATILRLMOGRVFEKIRIADGAIVTERDALIIGPLRV  
QFMYGVGVDDGRITLWRDYFDYDMFKGLRLVALVYPSLALTL"  
gene  
5355..6932  
/gene="PE\_PGRS"  
5355..6932  
/gene="PE\_PGRS"  
/note="RV2741, (MTV002.06), 1..: 525 aa; Member of M.  
tuberculosis PE\_PGRS subfamily, similar to many eg.

Query Match 98.9% Score 1220.8; DB 1: Length 56414;  
Best Local Similarity 99.8% Pred. No. 1.2e-166;  
Matches 123; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 ATCTTGAGATTATGCACTTTCTTCACTAGAGCTGATGATGAGAGGGTAATC 60  
DB 38201 ATCTTGAGATTATGCACTTTCTTCACTAGAGCTGATGATGAGAGGGTAATC 38260  
QY 61 ATGCGCGTCTGATTCGCAAGCAGCAAAAACAAGAAATTCGGGTCCATCACC 120  
DB 38261 ATGCGCGTCTGATTCGCAAGCAGCAAAAACAAGAAATTCGGGTCCATCACC 38320  
QY 121 GCCGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 180  
DB 38321 GCCGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 38380  
QY 181 GAGGCGTCTGCTATCACCAGCGGGATTTCAGAGCGGCGCAATGTCGGCAC 240  
DB 38381 GAGGCGTCTGCTATCACCAGCGGGATTTCAGAGCGGCGCAATGTCGGCAC 38440  
QY 241 GCCGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 300  
DB 38441 GCCGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 38500  
QY 301 GAATACGCGCGCTCTGCAAGCAGCGGCGGATTCAGAGCGGCGCAATGTCGGCAC 359  
DB 38501 GAATACGCGCGCTCTGCAAGCAGCGGCGGATTCAGAGCGGCGCAATGTCGGCAC 38560  
QY 360 CGTCTTCCACCATGCTGTTGGATTCCGCGCACACGTCATTCCTTACGAGACGCTC 419  
DB 38561 CGTCTTCCACCATGCTGTTGGATTCCGCGCACACGTCATTCCTTACGAGACGCTC 38620  
QY 420 CAGACGCGCGGAGCGCACTACCCCTGCTTGGCGGATGAGGAAATGCGCGGTGCAATC 479  
DB 38621 CAGACGCGCGGAGCGCACTACCCCTGCTTGGCGGATGAGGAAATGCGCGGTGCAATC 38680  
QY 480 GCCGCGCGGAGTGTGCGCTTACCACTGATGCGAACCCAAAGGCGGCGGTGCTGATG 539  
DB 38681 GCCGCGCGGAGTGTGCGCTTACCACTGATGCGAACCCAAAGGCGGCGGTGCTGATG 38740  
QY 540 GCGCGGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 599  
DB 38741 GCGCGGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 38800  
QY 600 TACAAAGCAGCGCGGATGCGCAAGCGGCGGATGAGTGTCTATCCAGGAGTCCGGA 659  
DB 38801 TACAAAGCAGCGCGGATGCGCAAGCGGCGGATGAGTGTCTATCCAGGAGTCCGGA 38860  
QY 660 ATGCGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 719  
DB 38861 ATGCGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 38920  
QY 720 TCGGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 779  
DB 38921 TCGGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 38980  
QY 780 GTGCGAGCGCGCAAGCAGCGGATGCGCAAGCGGCGGATGAGTGTCTATCCAGGAGTCCGGA 839  
DB 38981 GTGCGAGCGCGCAAGCAGCGGATGCGCAAGCGGCGGATGAGTGTCTATCCAGGAGTCCGGA 39040  
QY 840 GGTGCGAGTCTGAGGATATAGCATGACGAGGCGGCTTTTGAAGGCTGACGAGCG 899  
DB 39041 GGTGCGAGTCTGAGGATATAGCATGACGAGGCGGCTTTTGAAGGCTGACGAGCG 39100

QY 900 ACCACTACGACGACCGGAGCTTGGCGGTGACGACACGCTGTTTACTGCGTGGCAAC 959  
DB 39101 ACCACTACGACGACCGGAGCTTGGCGGTGACGACACGCTGTTTACTGCGTGGCAAC 39160  
QY 960 ATGCGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 1019  
DB 39161 ATGCGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 39220  
QY 1020 GTGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 1079  
DB 39221 GTGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 39280  
QY 1080 AAGGCTCTTTCGAGCGCAAGGAGGCGGCTTCTGCGGAGGCGGCGGAGTCCGGA 1139  
DB 39281 AAGGCTCTTTCGAGCGCAAGGAGGCGGCTTCTGCGGAGGCGGCGGAGTCCGGA 39340  
QY 1140 GTGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 1199  
DB 39341 GTGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 39400  
QY 1200 CACACGTCGCGGAGTAAAGGAGGAGTATGTCGGCC 1235  
DB 39401 CACACGTCGCGGAGTAAAGGAGGAGTATGTCGGCC 39436

## RESULT 12

LOCUS A87606 1235 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 3 from Patent WO9836089.  
ACCESSION A87606  
VERSION A87606.1 GI:5736246  
KEYWORDS

SOURCE unidentified.  
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 1235)  
Flobe/L. and Singh,M.  
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
FLOBE LEOPOLD (DE); SINGH MAHAVIR (DE)

FEATURES  
Source location/Qualifiers  
1..1235

BASE COUNT 236 a 394 c 382 g 220 t 3 others  
ORIGIN

Query Match 98.7% Score 1218.4; DB 5: Length 1235;  
Best Local Similarity 99.5% Pred. No. 6.4e-166;  
Matches 1230; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 ATCTTGAGATTATGCACTTTCTTCACTAGAGCTGATGATGAGAGGGTAATC 60  
DB 1 ATCTTGAGATTATGCACTTTCTTCACTAGAGCTGATGATGAGAGGGTAATC 60  
QY 61 ATGCGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 120  
DB 61 ATGCGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 120  
QY 121 GCGGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 180  
DB 121 GCGGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 180  
QY 181 GAGGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 240  
DB 181 GAGGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 240  
QY 241 GCGGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 300  
DB 241 GCGGCGTCTGCGGAACTAACCGGTGCGCATGAGTGTCTATCCAGGAGTCCGGA 300

```

QY 301 GAATACGGCCGCTGCGACACGGGC-GATCTTGTACGCTTCTTGCAATTTGGCCGGCTCA 359
DB 301 GAATACGGCCGCTGCGACACGGGCAGATCTTGTACGCTTCTTGCAATTTGGCCGGCTCA 360
QY 360 CGTGTTCACACGATCGCTTTGGATTCGGGACACGCTCAATTTGCCCTACAGACGCTC 419
DB 361 CGTGTTCACACGATCGCTTTGGATTCGGGACACGCTCAATTTGCCCTACAGACGCTC 420
QY 420 CAGACGGCCGAAAGGCGACATACCTGCTTGGCCGATGAGGAATGCGCGCTGACTC 479
DB 421 CAGACGGCCGAGCGGCGACATACCTGCTTGGCCGATGAGGAATGCGCGCTGACTC 480
QY 480 GCGCCAGGTTGGCGCTTACACGCTGATGCGAACCCAAAGGGGGCCGGTGTCTGATG 539
DB 481 GCGCCAGGTTGGCGCTTACACGCTGATGCGAACCCAAAGGGGGCCGGTGTCTGATG 540
QY 540 GGGGGGGTGGCGGGGCGAACGGGCGAGCGTGTGTGATGCGCGCGGACCGCGCGC 599
DB 541 GGGGGGGTGGCGGGGCGAACGGGCGAGCGTGTGTGATGCGCGCGGACCGCGCGC 600
QY 600 TACACGCAACCCGCACTCGCCAAAGCGATGGCGGCGACCGCTTACGCTTACACATCAAC 659
DB 601 TACACGCAACCCGCACTCGCCAAAGCGATGGCGGCGACCGCTTACGCTTACACATCAAC 660
QY 660 ATGACAAATTTGGGCACTCGACGCGGATGCTGGGGCGGATCGACACTGCTACTCA 719
DB 661 ATGACAAATTTGGGCACTCGACGCGGATGCTGGGGCGGATCGACACTGCTACTCA 720
QY 720 TCGGCTACGAGCTCGAGGCTGCGTCAACGTCGCGACCTGTGTGATGGGGCGCTCTG 779
DB 721 TCGGCTACGAGCTCGAGGCTGCGTCAACGTCGCGACCTGTGTGATGGGGCGCTCTG 780
QY 780 GTGCGACGGCCCAAGGACACCAATTAATGCTGCAATTAATTTGCCGATATGAACA 839
DB 781 GTGCGACGGCCCAAGGACACCAATTAATGCTGCAATTAATTTGCCGATATGAACA 840
QY 840 GTGCGAGTATGATATAGCATGACGACGAGGCGGCTTTGCAAGGCTCAGACG 899
DB 841 GTGCGAGTATGATATAGCATGACGACGAGGCGGCTTTGCAAGGCTCAGACG 900
QY 900 ACCACCTAGACACACCGGAGCTTGGCTTGCACGACAGCTGTTTACGCTGCGGAGAC 959
DB 901 ACCACCTAGACACACCGGAGCTTGGCTTGCACGACAGCTGTTTACGCTGCGGAGAC 960
QY 960 ATGCGCGCTCGGGGCGGAGAGCTGACCTAGCGGCTGACCAAGGAGAGATGCGCTAT 1019
DB 961 ATGCGCGCTCGGGGCGGAGAGCTGACCTAGCGGCTGACCAAGGAGAGATGCGCTAT 1020
QY 1020 GTGCTGAGCTTGGCAGCATGCTGGGCGGCGGCTGCGGCTGCAATCGGCACTAGCC 1079
DB 1021 GTGCTGAGCTTGGCAGCATGCTGGGCGGCGGCTGCGGCTGCAATCGGCACTAGCC 1080
QY 1080 AAGGCTTTTCAGCGACGAAAGGGGCTTACTGTCCGAAGGGGCGGCAACGCACTGGGG 1139
DB 1081 AAGGCTTTTCAGCGACGAAAGGGGCTTACTGTCCGAAGGGGCGGCAACGCACTGGGG 1140
QY 1140 GTGCGCTTACCGAGCCCGGCGAGGCTGCTGCTGACTCTCGGCGGCTCTTACGCGGAG 1199
DB 1141 GTGCGCTTACCGAGCCCGGCGAGGCTGCTGCTGACTCTCGGCGGCTCTTACGCGGAG 1200
QY 1200 CACACGTCGGAGTAAAGGAAGCATGATCGGC 1234
DB 1201 CACACGTCGGAGTAAAGGAAGCATGATCGGC 1235

```

```

RESULT 13
LOCUS AB9746 1235 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 3 from Patent W09832862.
ACCESSION AB9746
VERSION AB9746.1 GI:6738280
KEYWORDS
SOURCE unidentified.

```

```

ORGANISM unidentified
REFERENCE 1 (bases 1 to 1235)
AUTHORS Flohe L. and Singh M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHA VIR (DE)
FEATURES
source
1..1235
location/Qualifiers
BASE COUNT 236 a 394 c 382 g 220 t 3 others
ORIGIN
Query Match 98.7%; Score 1218.4; DB 5; Length 1235;
Best Local Similarity 99.6%; Pred. No. 6.4e-166;
Matches 1230; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
1 ATCTTGCAGATTAAATGCAATTTCTTCACTGAAGGCTACAGATGAGAGGGGTATC 60
DB 1 ATCTTGCAGATTAAATGCAATTTCTTCACTGAAGGCTACAGATGAGAGGGGTATC 60
QY 61 ATGCGCGTGGTATTCGACGAGACCAAAACAGAAATTCGGGTGGCATCACCCG 120
DB 61 ATGCGCGTGGTATTCGACGAGACCAAAACAGAAATTCGGGTGGCATCACCCG 120
QY 121 GCGGCGTGGGGAATACCCGCTGCGCATGAGTGTCTATCCAGGACAGTGGCGGA 180
DB 121 GCGGCGTGGGGAATACCCGCTGCGCATGAGTGTCTATCCAGGACAGTGGCGGA 180
QY 181 GAGGCGTGGGATACGAGCGGATTTCAAGGGGAGGCGGCACTGTGGGACC 240
DB 181 GAGGCGTGGGATACGAGCGGATTTCAAGGGGAGGCGGCACTGTGGGACC 240
QY 241 GCCGACAGGTGTGGGCGGAGCTGATTTATTTCTCAAGGTCMAAGAACGATAGCGCG 300
DB 241 GCCGACAGGTGTGGGCGGAGCTGATTTATTTCTCAAGGTCMAAGAACGATAGCGCG 300
QY 301 GAATACGGCGGCTCGACAGGCGG-GATCTTGTTCACGCTTCTTGCAATTTGGCCGGCTCA 359
DB 301 GAATACGGCGGCTCGACAGGCGGAGATCTTGTTCACGCTTCTTGCAATTTGGCCGGCTCA 360
QY 360 CGTGTTCACACGATCGCTTTGGATTCGGGACACGCTCAATTTGCCCTACAGACGCTC 419
DB 361 CGTGTTCACACGATCGCTTTGGATTCGGGACACGCTCAATTTGCCCTACAGACGCTC 420
QY 420 CAGACCGCGAAGCGCACTACCTGCTTGGCCGATGAGCGAAGTGGCGGCTGACTC 479
DB 421 CAGACCGCGAAGCGGCACTACCTGCTTGGCCGATGAGCGAAGTGGCGGCTGACTC 480
QY 480 GCGCGCAGTGGGGCTTACACGCTGATGCGAACCAAGGGGCGGCGGCTGTGATG 539
DB 481 GCGCGCAGTGGGGCTTACACGCTGATGCGAACCAAGGGGCGGCGGCTGTGATG 540
QY 540 GGGGGGGTGGCGGGGCGAACCGGCGGAGCTGTGTGATGAGGCGGCGGCGCGCGC 599
DB 541 GGGGGGGTGGCGGGGCGAACCGGCGGAGCTGTGTGATGAGGCGGCGGCGCGCGC 600
QY 600 TACACGCAACCCGCACTCGCCAAAGCGATGGCGGCGACCGCTTACGCTTACACATCAAC 659
DB 601 TACACGCAACCCGCACTCGCCAAAGCGATGGCGGCGACCGCTTACGCTTACACATCAAC 660
QY 660 ATGACAAATTTGGGCACTCGACGCGGATGCTGGGGCGGATCGACACTGCTACTCA 719
DB 661 ATGACAAATTTGGGCACTCGACGCGGATGCTGGGGCGGATCGACACTGCTACTCA 720
QY 720 TCGGCTACGAGCTCGAGGCTGCGTCAACGTCGCGACGCTGTGATTTGGGGCGCTCTG 779
DB 721 TCGGCTACGAGCTCGAGGCTGCGTCAACGTCGCGACGCTGTGATTTGGGGCGCTCTG 780
QY 780 GTGCGAGGCGGCAAGGCAACCAATTAATGCTGCAATTAATTTGCCGATATGAACA 839

```

Db 781 GTCGAGGCGCCAGGACCCAAATTAAGTCTGTCGCGCATATGAACCA 840  
 QY 840 GGTGGGATCTGTGTGATATACCATCGACGAGGCGGTGTTTGAAGGCTCAGACCG 899  
 Db 841 GGTGGGATCTGTGTGATATACCATCGACGAGGCGGTGTTTGAAGGCTCAGACCG 900  
 QY 900 ACCACCTACGACACCGGCTTGGCGGCGACGACGCTTTTACTGCGTGGGGAAC 959  
 Db 901 ACCACCTACGACACCGGCTTGGCGGCGACGACGCTTTTACTGCGTGGGGAAC 960  
 QY 960 ATGCCCGCTCGGTGGCGGAAGAGCTGACCTACGGCGTACCAAGCGAGATGCCGTAT 1019  
 Db 961 ATGCCCGCTCGGTGGCGGAAGAGCTGACCTACGGCGTACCAAGCGAGATGCCGTAT 1020  
 QY 1020 GTGCTCGACCTGCGGACCATGCGTGGCGGCGGTGCGGCTCGAATCCGCGCATAGCC 1079  
 Db 1021 GTGCTCGACCTGCGGACCATGCGTGGCGGCGGTGCGGCTCGAATCCGCGCATAGCC 1080  
 QY 1080 AAAGGCTCTTCGACGACGAGGCGGCTTACTGTCCTAACGGGTGCGACGACCTGGGG 1139  
 Db 1081 AAAGGCTCTTCGACGACGAGGCGGCTTACTGTCCTAACGGGTGCGACGACCTGGGG 1140  
 QY 1140 GTGCGCTTACCGAGCGCGCCAGCGTGTGCGCTGACTCTCGCGCGCTTACCGCGGAG 1199  
 Db 1141 GTGCGCTTACCGAGCGCGCCAGCGTGTGCGCTGACTCTCGCGCGCTTACCGCGGAG 1200  
 QY 1200 CACACGTGCGGAGTAAGGAGCGATGATGCGGC 1234  
 Db 1201 CACACGTGCGGAGTAAGGAGCGATGATGCGNC 1235

## RESULT 14

LOCUS A87608 1228 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 5 from Patent WO936089.  
 ACCESSION A87608  
 VERSION A87608.1 GI:5736248  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.

REFERENCE 1 (bases 1 to 1228)  
 Flohe, L. and Singh, M.  
 TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A-20-AUG-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)

FEATURES  
 source 1..1228  
 Location/Qualifiers  
 /organism="unidentified"  
 /db\_xref="taxon:32644"

BASE COUNT 236 a 391 c 382 g 219 t  
 ORIGIN

Query Match 98.2%; Score 1212.8; DB 5; Length 1228;  
 Best Local Similarity 99.8%; Pred. No. 4e-155;

Matches 1225; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 ATCTTGACATTAATGAATCTTCTTACACTGAAACCTACAGTATGAGAGGGTAAATC 60  
 Db 1 ATCTTGACATTAATGAATCTTCTTACACTGAAACCTACAGTATGAGAGGGTAAATC 60  
 QY 61 ATGCGGTGGGTATTCGAGCGGAGCAAAACAGAAATTCGGGGTGGGCAATCACCGCG 120  
 Db 61 ATGCGGTGGGTATTCGAGCGGAGCAAAACAGAAATTCGGGGTGGGCAATCACCGCG 120  
 QY 121 GCGGCGTGGGAACTAAACCGCTGTCGCAATGAGTCTCTCATCCAGGAGGTGCCGA 180  
 Db 121 GCGGCGTGGGAACTAAACCGCTGTCGCAATGAGTCTCTCATCCAGGAGGTGCCGA 180  
 QY 181 GAGGCTGGGCTATCACCGAGCGGATTCACAGGCGGCGGCACTGTGCGGACCC 240  
 Db 181 GAGGCTGGGCTATCACCGAGCGGATTCACAGGCGGCGGCACTGTGCGGACCC 240

QY 241 GCCGACAGGTGTGGCGCGACGCTGATTATTGCTCAAGTCGAAGAACCATAGCGCG 300  
 Db 241 GCCGACAGGTGTGGCGCGACGCTGATTATTGCTCAAGTCGAAGAACCATAGCGCG 300  
 QY 301 GAATAGGCGCGCTCGACACGAGCGG-GATCTGTTCAGCTTCTGATTTGGCGCGTCA 359  
 Db 301 GAATAGGCGCGCTCGACACGAGCGGAGATCTTGTTCAGCTTCTGATTTGGCGCGTCA 360  
 QY 360 GGTGCTTCCACCGATGCGTGTGATTCGCGACACGTCATTTGCTACGAGACGTC 419  
 Db 361 GGTGCTTCCACCGATGCGTGTGATTCGCGACACGTCATTTGCTACGAGACGTC 420  
 QY 420 CAGACGCGCGAAGGGCGCACTACCTGCTTGGCGGAGTGAAGTGGCGGCTGATC 479  
 Db 421 CAGACGCGCGAAGGGCGCACTACCTGCTTGGCGGAGTGAAGTGGCGGCTGATC 480  
 QY 480 GCCGCGCAGTTGGCGCTTACACCTGATGCGAACCCAAAGGGGCGCGGTGTGATG 539  
 Db 481 GCCGCGCAGTTGGCGCTTACACCTGATGCGAACCCAAAGGGGCGCGGTGTGATG 540  
 QY 540 GCGGCGGTGCGCGGTGCGAACCGCGCGACGTCGTGTGATGCGCGCGCGACCGCGGC 599  
 Db 541 GCGGCGGTGCGCGGTGCGAACCGCGCGACGTCGTGTGATGCGCGCGCGACCGCGGC 600  
 QY 600 TCAACGCGAGCGCGCATGCGCAAGGGCATGGCGGCGACGCTTACGAGTCAAC 659  
 Db 601 TCAACGCGAGCGCGCATGCGCAAGGGCATGGCGGCGACGCTTACGAGTCAAC 660  
 QY 660 ATCGACAAATCTGCGCACTCGACCGCGAGTCTGCGCGCGGATCGACACTCGTACTCA 719  
 Db 661 ATCGACAAATCTGCGCACTCGACCGCGAGTCTGCGCGCGGATCGACACTCGTACTCA 720  
 QY 720 TCGGCTTACGAGCTGAGGGTGCCTCAAAAGTGCAGCTGTGATTTGGGCGCTCTG 779  
 Db 721 TCGGCTTACGAGCTGAGGGTGCCTCAAAAGTGCAGCTGTGATTTGGGCGCTCTG 780  
 QY 780 GTGCGAGGCGCGCAAGGCAATTAAGTCTGATTAATTAATTAATTAATTAATTAATTA 839  
 Db 781 GTGCGAGGCGCGCAAGGCAATTAAGTCTGATTAATTAATTAATTAATTAATTAATTA 840  
 QY 840 GGTGGGATCTGTGTGATATACCATCGACGAGGCGGTGTTTGAAGGCTCAGACCG 899  
 Db 841 GGTGGGATCTGTGTGATATACCATCGACGAGGCGGTGTTTGAAGGCTCAGACCG 900  
 QY 900 ACCACCTACGACACCGCGCTTCCCGCTGCAGCAGACGCTTTTACTGCGTGGCGAAC 959  
 Db 901 ACCACCTACGACACCGCGCTTCCCGCTGCAGCAGACGCTTTTACTGCGTGGCGAAC 960  
 QY 960 ATGCCCGCTCGGTGGCGGAAGAGCTGACCTACGGCGTACCAAGCGAGATGCCGTAT 1019  
 Db 961 ATGCCCGCTCGGTGGCGGAAGAGCTGACCTACGGCGTACCAAGCGAGATGCCGTAT 1020  
 QY 1020 GTGCTCGACCTGCGGACCATGCGTGGCGGCGCGGTGCGGCTCGAATCGGCACTAGCC 1079  
 Db 1021 GTGCTCGACCTGCGGACCATGCGTGGCGGCGCGGTGCGGCTCGAATCGGCACTAGCC 1080  
 QY 1080 AAAGGCTTTCGACGACGAGGCGGTACTGTCCGAAGGGGTGCCACGACCTGGGG 1139  
 Db 1081 AAAGGCTTTCGACGACGAGGCGGTACTGTCCGAAGGGGTGCCACGACCTGGGG 1140  
 QY 1140 GTGCGCTTACCGAGCGCGCGAGCGTGTGCGCTGACTCTGCGCGCGCTTACCGCGGAG 1199  
 Db 1141 GTGCGCTTACCGAGCGCGCGAGCGTGTGCGCTGACTCTGCGCGCGCTTACCGCGGAG 1200  
 QY 1200 CACACGTGCGGAGTAAGGAGCGATGA 1227  
 Db 1201 CACACGTGCGGAGTAAGGAGCGATGA 1228

RESULT 15  
 LOCUS A89748 1228 bp DNA PAT 22-JAN-2000

DEFINITION Sequence 5 from Patent WO9832862.  
 ACCESSION A89748  
 VERSION A89748.1 GI:6738282  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1228)  
 AUTHORS Flohe L. and Singh M.  
 TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
 JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source 1..1228  
 Location/Qualifiers  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 236 a 391 c 382 g 219 t  
 ORIGIN

Query Match 98.2% Score 1212.8; DB 5; Length 1228;  
 Best Local Similarity 99.8%; Pred. No. 4e-165; 2; Indels 1; Gaps 1;  
 Matches 1225; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 ATCTTGAGATTATGCAACTTCTTTCACACTGAGGCTACGATGAGAGGGTAAATC 60  
 DB 1 ATCTTGAGATTATGCAACTTCTTTCACACTGAGGCTACGATGAGAGGGTAAATC 60  
 QY 61 ATCGGGGTGGTATTCGAGCGAGACCAAAAGCAATTCGGGGGCGCATACCCCG 120  
 DB 61 ATCGGGGTGGTATTCGAGCGAGACCAAAAGCAATTCGGGGGCGCATACCCCG 120  
 QY 121 GCCGGGCTCGGAGACTAACCCGTCGTGGCCATGAGTGCTCATCCAGGAGTGGCGA 180  
 DB 121 GCCGGGCTCGGAGACTAACCCGTCGTGGCCATGAGTGCTCATCCAGGAGTGGCGA 180  
 QY 181 GAGGGCTCGGCTATCACCGGCGGATTTCAAGCGGCGCGCACTGTGGGACC 240  
 DB 181 GAGGGCTCGGCTATCACCGGCGGATTTCAAGCGGCGCGCACTGTGGGACC 240  
 QY 241 GCCGACGAGTGTGGGCGGAGCTGATTTATGCTCAAGGCTCAAAAGACGATAGGGCG 300  
 DB 241 GCCGACGAGTGTGGGCGGAGCTGATTTATGCTCAAGGCTCAAAAGACGATAGGGCG 300  
 QY 301 GAATAGGCGCGCTGGGACACGGGC-GATCTGTTCACGTTCTTGTGATTTGGCGGCTCA 359  
 DB 301 GAATAGGCGCGCTGGGACACGGGC-GATCTGTTCACGTTCTTGTGATTTGGCGGCTCA 359  
 QY 360 CGTGTTCGACCATGCGTGTGTGATTCGGGACCAACGTCATTTGCTTACGAGACGCTC 419  
 DB 360 CGTGTTCGACCATGCGTGTGTGATTCGGGACCAACGTCATTTGCTTACGAGACGCTC 419  
 QY 420 CAGACGCGGAGGCGGACACTACCCGCTGTGGCGGATGAGCAAGTGGCGGCTGCACTC 479  
 DB 420 CAGACGCGGAGGCGGACACTACCCGCTGTGGCGGATGAGCAAGTGGCGGCTGCACTC 479  
 QY 481 GCCGCCAGGTTGGCGCTTACCACTGATGCAACCAAGGGGGCGCGGCTGTGCTGATG 539  
 DB 481 GCCGCCAGGTTGGCGCTTACCACTGATGCAACCAAGGGGGCGCGGCTGTGCTGATG 539  
 QY 540 GGGGGGTGGCGGCGGCTGCAACCGGCGGAGCTGTGTGATTCGGGCGCGGCGGCGGCG 599  
 DB 540 GGGGGGTGGCGGCGGCTGCAACCGGCGGAGCTGTGTGATTCGGGCGCGGCGGCGGCG 599  
 QY 600 TACAAGCGACCGCGCATGCGCAACGGGATGGGGCGGCTTACGGTTCTAGACATCAAC 659  
 DB 600 TACAAGCGACCGCGCATGCGCAACGGGATGGGGCGGCTTACGGTTCTAGACATCAAC 659  
 QY 660 ATGACAAATTTGGGCACTGACGCGGAGTTCTGGGCGGATCCCACTCGCTACTCA 719  
 DB 660 ATGACAAATTTGGGCACTGACGCGGAGTTCTGGGCGGATCCCACTCGCTACTCA 719  
 QY 720 TCGGCTAGAGCTCGAGGGTGGCGGCTCAAAAGTGGCGACCTGGTATGGGGCGGCTCTG 779  
 DB 720 TCGGCTAGAGCTCGAGGGTGGCGGCTCAAAAGTGGCGACCTGGTATGGGGCGGCTCTG 779

DB 721 TCGGCTAGAGCTCGAGGGTGGCGGCTCAAAAGTGGCGACCTGGTATGGGGCGGCTCTG 780  
 QY 780 GTGCGAGGGCGCAAGGCGACCAAAATTAGTCTGAAATTCATTTGTGCGCATATGAACCA 839  
 DB 781 GTGCGAGGGCGCAAGGCGACCAAAATTAGTCTGAAATTCATTTGTGCGCATATGAACCA 840  
 QY 840 GTGCGGCTACTGTGATTTAGCATGACGACGAGGGGCGTGTGTTGAAAGGCTCAGACCG 899  
 DB 841 GTGCGGCTACTGTGATTTAGCATGACGACGAGGGGCGTGTGTTGAAAGGCTCAGACCG 900  
 QY 900 ACCACCTAGACACCGCGGCTGCGCGGCGGACGACGACGCTGTTTACGCGTGGCGAGC 959  
 DB 901 ACCACCTAGACACCGCGGCTGCGCGGCGGACGACGACGCTGTTTACGCGTGGCGAGC 960  
 QY 960 ATGCGCGCTCGGTGCGGAGAGCTGACCTTACGCGGTGACCAACGCGAGATGCGGTAT 1019  
 DB 961 ATGCGCGCTCGGTGCGGAGAGCTGACCTTACGCGGTGACCAACGCGAGATGCGGTAT 1020  
 QY 1020 GTGCTGAGCTTGGCGGACCATGGCTGGCGGGGCGGTGCGGCTGCAATTCGGGCTAGCC 1079  
 DB 1021 GTGCTGAGCTTGGCGGACCATGGCTGGCGGGGCGGTGCGGCTGCAATTCGGGCTAGCC 1080  
 QY 1080 AAAGGCTTTTCGACGACGCAAGGGGCGTACTGTGCGAAAGGGGTGGCGACCGACTGGGG 1139  
 DB 1081 AAAGGCTTTTCGACGACGCAAGGGGCGTACTGTGCGAAAGGGGTGGCGACCGACTGGGG 1140  
 QY 1140 GTGCGCTTACCGGAGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1199  
 DB 1141 GTGCGCTTACCGGAGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
 QY 1200 CACACGTCGGAGTAAGGAGGAGCGATGA 1227  
 DB 1201 CACACGTCGGAGTAAGGAGGAGCGATGA 1228

Search completed: June 22, 2000, 14:56:13  
 Job time: 17788 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:10:02 ; Search time 458.59 Seconds  
(without alignments)  
673.777 Million cell updates/sec

Title: US-09-362-485-8

Perfect score: 1235  
Sequence: 1 ATCTTCGAGATTAATCGAAC.....GGGAAGCAGATGATCGGCC 1235

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1206.4	97.7	1245	1 V49626
2	1206.4	97.7	1260	1 V49510
3	1206.4	97.7	1260	1 V49525
4	454	36.8	682	1 V49511
5	265.2	21.5	1125	1 N91423
6	203.6	16.5	28171	1 V52155
7	166.4	13.5	9280	1 V74442
8	89.2	7.2	1074	1 T67971
9	69.2	5.6	344	1 V43039
10	56.8	4.6	31096	1 V74370
11	49.2	4.0	390	1 Q21833
12	49.2	4.0	390	1 Q36859
13	48.8	4.0	15872	1 T68715
14	46.4	3.8	985	1 V44439
15	46.4	3.8	985	1 V44439
16	45.4	3.7	833	1 V64568
17	45.4	3.7	5392	1 Q64203
18	45.4	3.7	5392	1 Q64201
19	44.8	3.6	17936	1 T93610
20	44.8	3.6	535	1 V44428
21	44.4	3.6	2582	1 V64537
22	44.4	3.6	17955	1 V56642
23	43.6	3.5	28958	1 T89566
24	43.6	3.5	28958	1 T89566
25	43.6	3.5	49377	1 V05287
26	42.8	3.5	114955	1 X53491
27	42.4	3.4	1833	1 O64206
28	42.4	3.4	2185	1 V84065
29	42.4	3.4	2186	1 Q89365
30	42.4	3.4	24379	1 T93095
31	42.4	3.4	24379	1 V25925
32	42.2	3.4	329	1 V44425
33	42.2	3.4	329	1 V64534
34	41.8	3.4	882	1 V44403

35	41.8	3.4	882	1 V64512	M. tuberculosis
36	41	3.3	3946	1 T93610	Mycobacterium tube
37	41	3.3	29879	1 Q46806	eyA region of S.
38	40.8	3.3	2151	1 Q76252	HSV-2 protease/IC
39	40.8	3.3	2151	1 Q76251	HSV-2 protease/ICP
40	40.8	3.3	2472	1 O84671	HSV-2 UL26 gene. N
41	40.8	3.3	20387	1 V62159	HSV-2 strain SB5 C
42	40.8	3.3	26338	1 V62134	HSV-2 strain SB5 C
43	40.8	3.3	117213	1 V62176	Infected cell prot
44	40.6	3.3	4257	1 V10362	The nucleotide seq
45	40.6	3.3	4257	1 V68520	

## ALIGNMENTS

RESULT 1	
V49626	V49626 standard; DNA; 1245 BP.
ID	
DT	20-NOV-1998 (first entry)
DE	Mycobacterium tuberculosis L-Alanine dehydrogenase 40 kD antigen.
KW	ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS	Mycobacterium tuberculosis.
PN	W09836089-A2.
PD	20-AUG-1998.
PF	29-JAN-1998; E00483.
PR	29-JAN-1997; EP-101338.
PA	(FLOH/) FLOHE L.
PI	Flohe L, Hutter B, Kolk A, Singh M;
DR	WPI: 98-4571.7/39.
PT	Diagnosis of, e.g., tuberculosis from alanine dehydrogenase activity
PT	- useful for, e.g., for diagnosis, differentiation of strains,
PT	monitoring vaccination and identification of mycobacterial
PT	inhibitors
PS	Disclosure, Fig 3.19, 55pp; German.
CC	The Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC	the production of kits for diagnosing tuberculosis (TB) and other
CC	mycobacterial infections in humans or animals. Kits are used for direct
CC	diagnosis of TB on clinical samples (e.g. body fluids) and can
CC	differentiate between pathogenic and non-virulent strains, e.g. for
CC	identifying pathogens of the M. tuberculosis (M. t.) complex. The kit may
CC	also be used to identify substances that inhibit mycobacteria, for
CC	combating epidemics and for vaccination follow-up. Oligonucleotides
CC	derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC	also for culture confirmation of isolated strains and for chromosome
CC	fingerprinting to detect/differentiate between mycobacteria, and for
CC	L-alanine-specific biotransformation reactions. AlaDH is an early
CC	antigen, present extracellularly after only a few days of growth, making
CC	it an ideal drug target.
SQ	Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 97.7%; Score 1206.4; DB 1; Length 1245;  
Best Local Similarity 99.4%; Pred. No. 8.1e-242;  
Matches 1234; Conservative 0; Mismatches 1; Indels 7; Gaps 2;

QY	1	ATCTTCGAGATTAATCGAAC	TTCTTCACTGACGTACAGTACAGTACGAGAGAGGATATC	60
DB	1	ATCTTCGAGATTAATCGAAC	TTCTTCACTGACGTACAGTACGAGAGAGGATATC	60
QY	61	ATGGGCGTGGTATTCGACG	AGACCAAAACAGG-----AATTCGGGTGGCATC	114
DB	61	ATGGGCGTGGTATTCGACG	AGACCAAAACAGG-----AATTCGGGTGGCATC	120
QY	115	ACCCGCGCGCGCTCGCGA	ACTAACCCGCTGCGCATGAGTGTCTATCCAGCAGGT	174
DB	121	ACCCGCGCGCGCTCGCGA	ACTAACCCGCTGCGCATGAGTGTCTATCCAGCAGGT	180
QY	175	GCCGAGAGGAGGCTGCTAT	CACCGACGCGGATTTCAAGCGCGCAGCGCAACTGGTC	234
DB	181	GCCGAGAGGAGGCTGCTAT	CACCGACGCGGATTTCAAGCGCGCAGCGCAACTGGTC	240



```

QY 235 GGCACCGCCGACACGAGTGTGGGGCGACGCTGATTATGCTCAAGGTCAAGAACCGATA 294
DB 241 GGCACCGCCGACACGAGTGTGGGGCGACGCTGATTATGCTCAAGGTCAAGAACCGATA 300
QY 295 GCGCGGAATATACGCGCGCTGCGACACGGGC -GATCTGTTCAGTCTTGCATTGGCC 353
DB 301 GCGCGGAATATACGCGCGCTGCGACACGGGCAGATCTTGTTCAGTCTTGCATTGGCC 360
QY 354 GCGTACAGTCTTGCACACCATGCTTGTGATTCCGGACACGCTCAANTGGCTACGAG 413
DB 361 GCGTACAGTCTTGCACACCATGCTTGTGATTCCGGACACGCTCAANTGGCTACGAG 420
QY 414 ACCGTACAGCGCGCGGACGACGACGACGCTGCTGCGGATGAGCAAGTGGCGGT 473
DB 421 ACCGTACAGCGCGCGGACGACGACGACGCTGCTGCGGATGAGCAAGTGGCGGT 480
QY 474 CGACTCGCGCCGACGAGTGGCGCTTACCGCTGATGCGAACCCAGGGGGCGCGGTG 533
DB 481 CGACTCGCGCCGACGAGTGGCGCTTACCGCTGATGCGAACCCAGGGGGCGCGGTG 540
QY 534 CTGATGGGGGGGGGGCGCGGGGTGGAACCGGACGCTGCTGATGCGCGCGCGAC 593
DB 541 CTGATGGGGGGGGGGCGCGGGGTGGAACCGGACGCTGCTGATGCGCGCGCGAC 600
QY 594 GCGCGTACAAACGACGCGCGGATGCGACGCGATGGGGCGGACGCTTACGCTTAC 653
DB 601 GCGCGTACAAACGACGCGCGGATGCGACGCGGATGGGGCGGACGCTTACGCTTAC 660
QY 654 ATCAACATCGAACAAATCTGGCACTGAGCGCGGATGCTGCGCGCGGATGCGAC 713
DB 661 ATCAACATCGAACAAATCTGGCACTGAGCGCGGATGCTGCGCGCGGATGCGAC 720
QY 714 TACTCATCGGCTTACGAGCTGAGGGTGGCGCTCAACGCTGCGGATGCTGAGGG 773
DB 721 TACTCATCGGCTTACGAGCTGAGGGTGGCGCTCAACGCTGCGGATGCTGAGGG 780
QY 774 GTCTGTGGGGGGGGCGGCAAGGCAACCAATTAATGCTGCAATGCTGCGGATG 833
DB 781 GTCTGTGGGGGGGGCGGCAAGGCAACCAATTAATGCTGCAATGCTGCGGATG 840
QY 834 AAACGAGTGGCTACTGCTGATATACCATGACACGAGGGGCTGTTTGAAGCTCA 893
DB 841 AAACGAGTGGCTACTGCTGATATACCATGACACGAGGGGCTGTTTGAAGCTCA 900
QY 894 CGACCGACGACCTACGACACCGACGCTGCGGATGCGACGAGCTGTTTACGCTG 953
DB 901 CGACCGACGACCTACGACACCGACGCTGCGGATGCGACGAGCTGTTTACGCTG 960
QY 954 GCGAATGCTGCGCTGCTGCGGACGCTGACGCTACGCGCTGACCAACGCGAGATG 1013
DB 961 GCGAATGCTGCGCTGCTGCGGACGCTGACGCTACGCGCTGACCAACGCGAGATG 1020
QY 1014 CCGTATGCTGCTGAGCTTGGCGACCATGGCTGGGGGGGGGGCTGCAATCCGGCA 1073
DB 1021 CCGTATGCTGCTGAGCTTGGCGACCATGGCTGGGGGGGGGGCTGCAATCCGGCA 1080
QY 1074 CTAGCGAAGGCTTTCGACGACGAGGCGCTTACTGCTGCAAGGAGGTGGCCAGC 1133
DB 1081 CTAGCGAAGGCTTTCGACGACGAGGCGCTTACTGCTGCAAGGAGGTGGCCAGC 1140
QY 1134 CTGGGGGTGCGCTTACGAGCGCGCGACGCTGCTGCTGCTGCTGCTGCTGCT 1193
DB 1141 CTGGGGGTGCGCTTACGAGCGCGCGCGACGCTGCTGCTGCTGCTGCTGCTGCT 1200
QY 1194 GCGGAGCAGCTGCGGAGTAAAGGAGGAGGATGATGCGGC 1235
DB 1201 GCGGAGCAGCTGCGGAGTAAAGGAGGAGGATGATGCGGC 1242

```

RESULT 2  
ID V49510 standard; DNA: 1260 BP.  
AC V49510;

DE 20-Oct-1998 (first entry)  
DE Mycobacterium sp. AladH DNA.  
KW Alanine dehydrogenase; AladH; ADH: diagnosis; tuberculosis; pathogen;  
KW swimmers disease; vaccine; epidemic; infection; identification; ss.  
OS Mycobacterium sp.  
PN WO9832862-A2.  
PD 30-JUL-1998.  
PF 29-JAN-1998; E00484.  
PR 29-JAN-1997; EP-101339.  
PA (FLOH/) FLOHE L.  
PI FLOHE L., Huter B., Kolk A., Singh M.  
DR WPI: 98-427958/36.  
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum  
PT - used for diagnosis of tuberculosis and other mycobacterial  
PT diseases, also for treatment and prevention, for drug screening and  
PT for bio-transformation.  
PS Disclosure: Page 11: 57pp; German.  
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated  
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis  
CC and other mycobacterial infections (including 'swimmers' disease', caused  
CC by M. marinum, a fish pathogen) in humans or animals. The protein can  
CC also be used for control of epidemics and for vaccination, to screen for  
CC agents with anti-mycobacterial activity, and in bio-transformations that  
CC are specific for L-alanine. Also mycobacteria can be identified by  
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted  
CC early during infection.  
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 97.7%; Score 1206.4; DB 1; Length 1260;  
Best Local Similarity 99.4%; Pred. No. 8.1e-24;  
Matches 1234; Conservative 0; Mismatches 1; Indels 7; Gaps 2;

```

QY 1 ATCTTCAATTAATGAATCTTCTTACAGCTGACGATGACGATGAGGGTAAATC 60
DB 16 ATCTTCAATTAATGAATCTTCTTACAGCTGACGATGACGATGAGGGTAAATC 75
QY 61 ATGCGGCTGGTATTCGACCGGACCAAAACAGC-----AATTCGGGTGGCCATC 114
DB 76 ATGCGGCTGGTATTCGACCGGACCAAAACAGC-----AATTCGGGTGGCCATC 135
QY 115 ACCCGGCGCGCTCGCGGAATACCGCTGTGGCATGAGGTGCTATCCAGGAGGT 174
DB 136 ACCCGGCGCGCTCGCGGAATACCGCTGTGGCATGAGGTGCTATCCAGGAGGT 195
QY 175 GCGGAGAGGGCTGCTGCTATACCGACCGGATTTAAGCGGCGGACCGCACTGCTC 234
DB 196 GCGGAGAGGGCTGCTGCTATACCGACCGGATTTAAGCGGCGGACCGCACTGCTC 255
QY 235 GGCACCGCCGACGAGTGTGGGGCGACGCTGATTATTGCTCAAGGTCAAGAACCGATA 294
DB 256 GGCACCGCCGACGAGTGTGGGGCGACGCTGATTATTGCTCAAGGTCAAGAACCGATA 315
QY 295 GCGGCGGAATACGCGCGCTGCGACACGGGC -GATCTGTTCAGTCTTGCATTGGCC 353
DB 316 GCGGCGGAATACGCGCGCTGCGACACGGGCAGATCTTGTTCAGTCTTGCATTGGCC 375
QY 354 GCGTACAGTCTTGCACACCATGCTTGTGATTCCGGACACGCTCAANTGGCTACGAG 413
DB 376 GCGTACAGTCTTGCACACCATGCTTGTGATTCCGGACACGCTCAANTGGCTACGAG 435
QY 414 ACCGTACAGCGCGCGGACGACGACGACGCTGCTGCGGATGAGCAAGTGGCGGT 473
DB 436 ACCGTACAGCGCGCGGACGACGACGACGCTGCTGCGGATGAGCAAGTGGCGGT 495
QY 474 CGACTCGCGCCGACGAGTGGCGCTTACCGCTGATGCGAACCCAGGGGGCGCGGTG 533
DB 496 CGACTCGCGCCGACGAGTGGCGCTTACCGCTGATGCGAACCCAGGGGGCGCGGTG 555
QY 534 CTGATGGGGGGGGGGCGCGGGGTGGAACCGGACGCTGCTGATGCGCGCGGAC 598
DB 556 CTGATGGGGGGGGGGCGCGGGGTGGAACCGGACGCTGCTGATGCGCGCGGAC 615

```

CC derived from AladH are used similarly in diagnostic hybridisation tests, CC also for culture confirmation of isolated strains and for chromosome fingerprinting to detect/differentiate between mycobacteria, and for CC L-alanine-specific biotransformation reactions. AladH is an early CC antigen, present extracellularly after only a few days of growth, making it an ideal drug target.

Sequence 1260 bp; 243 A; 403 C; 389 G; 225 T;

Query Match 97.7%; Score 1206.4; DB 1; Length 1260; Best Local Similarity 99.4%; Pred. No. 8.1e-242; Matches 1234; Conservative 0; Mismatches 1; Indels 7; Gaps 2;

1 ATCTGAGATTATGAGCTTCTTCTACACTGAGCGTACAGTATGAGAGGATGATC 60  
 16 ATCTTGACGATTAAATGACATTCTTCTACACTGAGCGTACAGTATGAGAGGATGATC 75  
 61 ATGCGGCTGCTATTCGACCGAGACCAAAACACAG-----AATTCGGGTGGCCATC 114  
 76 ATGCGGCTGCTATTCGACCGAGACCAAAACACAGATTCCATTCCGGGTGGCCATC 135  
 115 ACCCGCGCGCGCTCGCGGACCTAACCCGTCGCGCATGAGGTGCTATCCAGGATG 174  
 136 ACCCGCGCGCGCTCGCGGACCTAACCCGTCGCGCATGAGGTGCTATCCAGGATG 195  
 175 GCGGAGAGGGGCTCGGCTATACCGAGCGGATTCAGGCGGACAGCGCAACTGTC 234  
 196 GCGGAGAGGGGCTCGGCTATACCGAGCGGATTCAGGCGGACAGCGCAACTGTC 255  
 235 GGCAGCGCGGACGAGTGTGGGCGGACGCTGATTATTGCTCAAGTCAAGAACGAT 294  
 256 GGCAGCGCGGACGAGTGTGGGCGGACGCTGATTATTGCTCAAGTCAAGAACGAT 315  
 295 GCGCGGAATAGGCGCGCTCGGACACGCGGACGCTGATTATTGCTCAAGTCAAGAACGAT 353  
 316 GCGCGGAATAGGCGCGCTCGGACACGCGGACGCTGATTATTGCTCAAGTCAAGAACGAT 375  
 354 GCGTCAAGTGC.TGCAACGATGCGTGTGGATTCGCGGACGCTCAATGGCTACGAG 413  
 376 GCGTCAAGTGC.TGCAACGATGCGTGTGGATTCGCGGACGCTCAATGGCTACGAG 435  
 414 ACCGTCAAGCGCGCGGACGCTGAGTGTGGATTCGCGGACGCTCAATGGCTACGAG 473  
 436 ACCGTCAAGCGCGCGGACGCTGAGTGTGGATTCGCGGACGCTCAATGGCTACGAG 495  
 474 CGACTCGCGCGCGGACGCTGAGTGTGGATTCGCGGACGCTCAATGGCTACGAG 533  
 496 CGACTCGCGCGCGGACGCTGAGTGTGGATTCGCGGACGCTCAATGGCTACGAG 555  
 534 CTGATGGGCGGGGTGCCCGGCTGCAACCGGCGGACGCTGAGTGTGGATTCGCGGACG 593  
 556 CTGATGGGCGGGGTGCCCGGCTGCAACCGGCGGACGCTGAGTGTGGATTCGCGGACG 615  
 594 GCGGCGTACAAAGCAGCGCGCATCGCAACGGATGGGCGGACGCTGAGTGTGGATTCGAG 653  
 616 GCGGCGTACAAAGCAGCGCGCATCGCAACGGATGGGCGGACGCTGAGTGTGGATTCGAG 675  
 654 ATCAACATCGACAAACTTTCGCAACTCGAGCGGATTCGCGGCGGATTCGCACTCGC 713  
 676 ATCAACATCGACAAACTTTCGCAACTCGAGCGGATTCGCGGCGGATTCGCACTCGC 735  
 714 TACTCATCGGCTTACGAGCTCGAGGCGGCTGCAACGTCGCACTGAGTGGGCGC 773  
 736 TACTCATCGGCTTACGAGCTCGAGGCGGCTGCAACGTCGCACTGAGTGGGCGC 795  
 774 GTCTGTGTCAGGCGGCAAGGACCAAAATGATTCGCACTGAGTGGGCGCATATG 833  
 796 GTCTGTGTCAGGCGGCAAGGACCAAAATGATTCGCACTGAGTGGGCGCATATG 855  
 834 AAACCAAGTCCGTACTGCTGATATAGCATATGACAGGCGGCTGTTGCAAGGCTCA 893  
 856 AAACCAAGTCCGTACTGCTGATATAGCATATGACAGGCGGCTGTTGCAAGGCTCA 915

CC derived from AladH are used similarly in diagnostic hybridisation tests, CC also for culture confirmation of isolated strains and for chromosome fingerprinting to detect/differentiate between mycobacteria, and for CC L-alanine-specific biotransformation reactions. AladH is an early CC antigen, present extracellularly after only a few days of growth, making it an ideal drug target.

Sequence 1260 bp; 243 A; 403 C; 389 G; 225 T;

Query Match 97.7%; Score 1206.4; DB 1; Length 1260; Best Local Similarity 99.4%; Pred. No. 8.1e-242; Matches 1234; Conservative 0; Mismatches 1; Indels 7; Gaps 2;

1 ATCTGAGATTATGAGCTTCTTCTACACTGAGCGTACAGTATGAGAGGATGATC 60  
 16 ATCTTGACGATTAAATGACATTCTTCTACACTGAGCGTACAGTATGAGAGGATGATC 75  
 61 ATGCGGCTGCTATTCGACCGAGACCAAAACACAG-----AATTCGGGTGGCCATC 114  
 76 ATGCGGCTGCTATTCGACCGAGACCAAAACACAGATTCCATTCCGGGTGGCCATC 135  
 115 ACCCGCGCGCGCTCGCGGACCTAACCCGTCGCGCATGAGGTGCTATCCAGGATG 174  
 136 ACCCGCGCGCGCTCGCGGACCTAACCCGTCGCGCATGAGGTGCTATCCAGGATG 195  
 175 GCGGAGAGGGGCTCGGCTATACCGAGCGGATTCAGGCGGACAGCGCAACTGTC 234  
 196 GCGGAGAGGGGCTCGGCTATACCGAGCGGATTCAGGCGGACAGCGCAACTGTC 255  
 235 GGCAGCGCGGACGAGTGTGGGCGGACGCTGATTATTGCTCAAGTCAAGAACGAT 294  
 256 GGCAGCGCGGACGAGTGTGGGCGGACGCTGATTATTGCTCAAGTCAAGAACGAT 315  
 295 GCGCGGAATAGGCGCGCTCGGACACGCGGACGCTGATTATTGCTCAAGTCAAGAACGAT 353  
 316 GCGCGGAATAGGCGCGCTCGGACACGCGGACGCTGATTATTGCTCAAGTCAAGAACGAT 375  
 354 GCGTCAAGTGC.TGCAACGATGCGTGTGGATTCGCGGACGCTCAATGGCTACGAG 413  
 376 GCGTCAAGTGC.TGCAACGATGCGTGTGGATTCGCGGACGCTCAATGGCTACGAG 435  
 414 ACCGTCAAGCGCGCGGACGCTGAGTGTGGATTCGCGGACGCTCAATGGCTACGAG 473  
 436 ACCGTCAAGCGCGCGGACGCTGAGTGTGGATTCGCGGACGCTCAATGGCTACGAG 495  
 474 CGACTCGCGCGCGGACGCTGAGTGTGGATTCGCGGACGCTCAATGGCTACGAG 533  
 496 CGACTCGCGCGCGGACGCTGAGTGTGGATTCGCGGACGCTCAATGGCTACGAG 555  
 534 CTGATGGGCGGGGTGCCCGGCTGCAACCGGCGGACGCTGAGTGTGGATTCGCGGACG 593  
 556 CTGATGGGCGGGGTGCCCGGCTGCAACCGGCGGACGCTGAGTGTGGATTCGCGGACG 615  
 594 GCGGCGTACAAAGCAGCGCGCATCGCAACGGATGGGCGGACGCTGAGTGTGGATTCGAG 653  
 616 GCGGCGTACAAAGCAGCGCGCATCGCAACGGATGGGCGGACGCTGAGTGTGGATTCGAG 675  
 654 ATCAACATCGACAAACTTTCGCAACTCGAGCGGATTCGCGGCGGATTCGCACTCGC 713  
 676 ATCAACATCGACAAACTTTCGCAACTCGAGCGGATTCGCGGCGGATTCGCACTCGC 735  
 714 TACTCATCGGCTTACGAGCTCGAGGCGGCTGCAACGTCGCACTGAGTGGGCGC 773  
 736 TACTCATCGGCTTACGAGCTCGAGGCGGCTGCAACGTCGCACTGAGTGGGCGC 795  
 774 GTCTGTGTCAGGCGGCAAGGACCAAAATGATTCGCACTGAGTGGGCGCATATG 833  
 796 GTCTGTGTCAGGCGGCAAGGACCAAAATGATTCGCACTGAGTGGGCGCATATG 855  
 834 AAACCAAGTCCGTACTGCTGATATAGCATATGACAGGCGGCTGTTGCAAGGCTCA 893  
 856 AAACCAAGTCCGTACTGCTGATATAGCATATGACAGGCGGCTGTTGCAAGGCTCA 915

QY 894 CGACGACACCTACGACACACCGACGCTCCGCTGACACGACGCTTTTACTGCGTG 953  
 |||  
 DB 916 CGACGACACCTACGACACACCGACGCTCCGCTGACACGACGCTTTTACTGCGTG 975  
 QY 954 GCGAATATGCGCGCTGCTGCGACGACGCTGACGCTGACGCTGACGACGATG 1013  
 |||  
 DB 976 GCGAATATGCGCGCTGCTGCGACGACGCTGACGCTGACGCTGACGACGATG 1035  
 QY 1014 CCGATATGCTGACGCTGCGACGACGCTGCGCGCGCGCTGCGGATGCGGCA 1073  
 |||  
 DB 1036 CCGATATGCTGACGCTGCGACGACGCTGCGCGCGCGCTGCGGATGCGGCA 1095  
 QY 1074 CTACGCAAAAGCTTTTGACGACGACGAGGCGCTTACTGCTGCAAGGCTGCGACGAC 1133  
 |||  
 DB 1096 CTACGCAAAAGCTTTTGACGACGACGAGGCGCTTACTGCTGCAAGGCTGCGACGAC 1155  
 QY 1134 CTGGGGGCTGCGCTTACGACGCGCGACGCTGCTGCTGCTGCTGCTGCTGCTTAC 1193  
 |||  
 DB 1156 CTGGGGGCTGCGCTTACGACGCGCGACGCTGCTGCTGCTGCTGCTGCTGCTTAC 1215  
 QY 1194 GCGGACACACGCTGCGGAGTAAAGGAGCGATGATGCTGCGC 1235  
 |||  
 DB 1216 GCGGACACACGCTGCGGAGTAAAGGAGCGATGATGCTGCGC 1257

## RESULT 4

VA9511  
 ID V49511 standard; DNA; 682 BP.  
 AC V49511:  
 DE 20-OCT-1998 (first entry)  
 KW Mycobacterium marinum Mar3 DNA.  
 KW Alanine dehydrogenase; AlADH; ADH; diagnosis; tuberculosis; pathogen;  
 KW swimmers disease; vaccine; epidemic; infection; identification; ss.  
 OS Mycobacterium murinum.  
 FH Key location/Qualifiers  
 FT CDS 1..682  
 FT /\*tag= a  
 FT /codon\_start= 3  
 FT /product= "Mar3"  
 FT /note= "Alanine dehydrogenase"  
 PN M09832862-A2.  
 PD 30-JUL-1998.  
 PF 29-JAN-1998; E00484.  
 PR 29-JAN-1997; EP-101339.  
 PA (FLOH/) FLOHE L., Kolk A., Singh M;  
 PI Flohe L., Hutter B.  
 DR MPI: 98-427958/36.  
 DR P-PSDB: W64481.  
 PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum  
 PT - used for diagnosis of tuberculosis and other mycobacterial  
 PT diseases; also for treatment and prevention, for drug screening and  
 PT for bio-transformation  
 PS Claim 1: Page 34-35; 57pp; German.  
 CC This sequence encodes an alanine dehydrogenase (ADH) protein, Mar3  
 CC isolated from Mycobacterium murinum. This protein is used to diagnose  
 CC tuberculosis and other mycobacterial infections (including 'swimmers'  
 CC disease', caused by *M. marinum*, a fish pathogen) in humans or animals.  
 CC The protein can also be used for control of epidemics and for  
 CC vaccination, to screen for agents with anti-mycobacterial activity, and  
 CC in bio-transformations that are specific for L-alanine. Also mycobacteria  
 CC can be identified by analysis of genomic ADH sequences. ADH is an antigen  
 CC that is secreted early during infection.  
 SO Sequence 682 BP; 105 A; 254 C; 225 G; 98 T;

Query Match 36.8%; Score 454; DB 1; Length 682;  
 Best Local Similarity 80.1%; Pred. No. 5.4e-86;  
 Matches 546; Conservative 0; Mismatches 135; Indels 1; Gaps 1;

QY 98 AATTCGCGGTACGACACCGCGCGCTGCGGAACTAACCGCTGCGGCAATGAGG 157  
 |||  
 DB 1 AATTCGCGGTACGACACCGCGCGCTGCGGAACTAACCGCTGCGGCAATGAGG 60

QY 158 TGTGATCCAGAGGAGTGCCTGCGAGAGGCTCGGCTATTCACGACGCGGATTCAGAGCGG 217  
 |||  
 DB 61 TGTGATCCAGAGGAGTGCCTGCGAGAGGCTCGGCTATTCACGACGCGGATTCAGAGCGG 120  
 QY 218 CAGGCGCGCAATGCTGCTGCGACGCGCGACGAGTGTGGCGCGACGCTGATTTATGCTCA 277  
 |||  
 DB 121 CCGGATCCAGCTGATACGACGCGCA CAGGCTGCGGAGTGTGGCGCGACGCTGCTCA 180  
 QY 278 AGGTCAAGAGACGATAGCGCGGATACGCGCGCTGCGACAGGCGG -GATCTGTTCA 336  
 |||  
 DB 181 AGGTCAAGAGACGATAGCGCGGATACGCGCGCTGCGACAGGCGG -GATCTGTTCA 240  
 QY 337 CGTTCTGCAATTTGGCGCGCTACGCTGCTTGAACGAGTGGTTGATTCGCGACCA 396  
 |||  
 DB 241 CCGTACCTGCACTGGCGCGCTGCGCGCTGCGACGATGCTGCTGATGAGTCTGCGACCA 300  
 QY 397 CGTCAATGCTGCTGCGACGCTGCGACGCGCGGAGGCGGCTATACCCCTGCTTGGCCCA 456  
 |||  
 DB 301 CGTCAATGCTGCTGCGACGCTGCGACGCGCGGAGGCGGCTATACCCCTGCTTGGCCCA 360  
 QY 457 TGAGCAAGTCCGCGGCTGCGACTGCGCGCGGAGTGGCGCTTACGACCTGATGCGAACCC 516  
 |||  
 DB 361 TGAGCAAGTCCGCGGCTGCGACTGCGCGCGGAGTGGCGCTTACGACCTGATGCGAACCC 420  
 QY 517 AAGGCGCGCGGCTGCTGATGCGGCGGCTGCGCGGCTGCGACCTGCGGCTGCGGCTGCG 576  
 |||  
 DB 421 ACGGCGGCTGCGGCTGCTGATGCGGCGGCTGCGCGGCTGCGGCTGCGGCTGCGGCTGCG 480  
 QY 577 TGATCGGCGCGCGGCTGCGACTGCGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 636  
 |||  
 DB 481 TGATCGGCGCGCGGCTGCGACTGCGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 540  
 QY 637 CGTTACGCTTACGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 696  
 |||  
 DB 541 TGGTCACTGCTGATGCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 600  
 QY 697 GCGGATCCACACTGCTGATGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 756  
 |||  
 DB 601 GCGGCTGCGGAGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 660  
 QY 757 ACTGTGATTTGGGCGCGCTCT 778  
 |||  
 DB 661 ACATGTGATGCGGCGCGCTCT 682

## RESULT 5

N91423  
 ID N91423 standard; DNA; 1125 bp.  
 AC N91423:  
 DE 01-FEB-1991 (first entry)  
 DE Sequence of heat-resistant alanine dehydrogenase (AH) gene with mol. wt.  
 DE 2 MD or less  
 KW Enzyme: ds.  
 OS Bacillus stearothermophilus IFO 12550.  
 PN J01043194-A.  
 PD 15-FEB-1989.  
 PF 10-AUG-1987; 200524.  
 PR 10-AUG-1987; JP-200524.  
 PA (NIRA) Unilika KK.  
 DR MPI: 89-096096/13.  
 PT Recombinant plasmid for transforming Escherichia coli -  
 PT obt. by connecting heat resistant alanine dehydrogenase gene to  
 PT vector plasmid  
 PS Disclosure: Fig 2, p693; 9pp; Japanese.  
 CC A recombinant plasmid contg. heat-resistant alanine dehydrogenase (AH)  
 CC gene with mol. wt. 2 or less MD is claimed. Cells transformed with the  
 CC vector produce high levels of heat-resistant AH. Transformed E. coli  
 CC is useful as a clinical inspection reagent.  
 SO Sequence 1125 BP; 246 A; 295 C; 367 G; 217 T;

Query Match 21.5%; Score 265.2; DB 1; Length 1125;  
 Best Local Similarity 56.1%; Pred. No. 7.1e-47;

Matches	603:	Conservative	0:	Mismatches	458:	Indels	13:	Gaps	5:
QY	61	ATGCCGCTGATTTTCGACCGAGACCAAAACACGAATTCGGGTGGCCATCACCCG	120						
Db	1	ATGAAAGATGGGATTCACAAAGAAATCAAAACAAATGAAGAAACCGCGTCCCATCTCCG	60						
QY	121	GGCGGCTGGGGAATACACCGCTGCTGGCATGAGTGTCT---CATCCAGCAGGTGTC	177						
Db	61	GAGGCGGTATGACGCTGCTCAAGCGGGGACATGAGTGTGTAA.GTGAAGACGGAAGG	120						
QY	178	GAGAGAGGCTCGATATCACCGACGCGGATTTCAAGCGCGACGCGCATGTGTGCG	237						
Db	121	GGCGCTGGGCTGGGGTTTCCGATTCGAGTATGAAGAAACCGGGGACACATGATGCTTC	180						
QY	238	ACCGCGACAGTGTGGGCGGACGCTGATTTATTTCTCAAGTCAAAACACCATATGCG	297						
Db	181	CGAAGCTGGAGAGATGCTTGGACGGGGAGATGTGTAAAGTGAAGAGCGCGTGGCT	240						
QY	298	GCGAATACGGCGCGCTGCGACACG-GCGATCTTGTACGCTTTCGATTGGCGCG	356						
Db	241	CGAGAGTTCGGTATTTTGGCCCCGATGATTTTTCATTTTTCATTTTACCGCG	300						
QY	357	TCACGCTGCTGACCGATCGCTTTGATTCGGCACACGTCATTTGCTTACGAGAC	416						
Db	301	GCCGAACGGCTACGAAACCGCTGCTGAGCAAAAGTGTGGCATCGCTTACGAGAC	360						
QY	417	GTCACAGACCGGAGGCGGACATACCCCTGCTGGCCGATGAGGAATCGCGCGTGA	476						
Db	361	GTCGACGCTGCGAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	420						
QY	477	CTGCGCGCGGAGTGTGGCTTACACATGATGCAACCCCAAGGGGCGCGGTGCTG	536						
Db	421	ATGTGCTGCAATGCGCGCCAGTTTCTGAGAGCCGACGGGGGAAAGGCGATTTTG	480						
QY	537	ATGGGCGGGTCCCGCGCTGCAACCGCGGACGCTGCTGATGCGCGCGGACCGCC	596						
Db	481	CTGGCGGGTCCCGGAGTGTGGGCGGCGCAAGTACATATGCGCGCGGCAACGCG	540						
QY	597	GCGTCAACAGCGCCGATCGCAACGCGCATGGCGGCGCGGATTCAGGTTTACATC	656						
Db	541	GCGAGAAAGCGGGGAAATATCGGTGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTG	600						
QY	657	AACATCGAACAATTCGCGACCTGACCGCGGATTCGCGGCGGATTCACATGCTAC	716						
Db	601	AACGCGGCGGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	660						
QY	717	TCATGCGCTTACGAGCTGAGGCTGCGCAACGCTGCGACCTGCTGATTTGGGCGCTC	776						
Db	661	TCCAACTGTAACATATGCGCGAGTGTGCGGCAATCGGATTTGTGCTGCGCTGCTC	720						
QY	777	CTGCTGCGAGGCGCCAGGACCCAAATTAATTCGAAATTCATTTGCGCATATGA	836						
Db	721	TTGATCCCGGGGCGAA--AGCGAAGCTGTGTGAGGAAAGATGTGCTGCTGCTGCTG	777						
QY	837	CGAGGTGGGATCTGATGATATACCATTCGACGCGGCGGCTGTTTGA--GCTCA	893						
Db	778	CGGGATGGGTGTGTGTATGCGCATTCGACCAAGCGCGCATTTTGAAGACGCCAC	837						
QY	894	CGACGACCACTTACGACACCGAGCTGCGCGTGCAGACGACGCTTTTATGAGGCTG	953						
Db	838	CGACGCTGAGAGAGATGCGCATATGCTCAAGACGCGGTGTGCTCATTTAGCCCTC	897						
QY	954	GCGAACAATGCGCGCTGCGTGGGAGAGCTGAGCTTACGCGGCTGACCAAGCGAGTG	1013						
Db	898	GCCAAATTTGGCGGCG--GTGCGCGGACGTCGACATTCGCGCTTACCAAGCTGCATC	954						
QY	1014	CGGATGTGCTGAGCTTGGACGACATGCTGGGCGGGGCTGCGGCTGCAATCGGCA	1073						
Db	955	CCATACGCTTGCATATGCGCAACAAGGCTACCGCGCGGTGCTTGGATTAACCGCGCG	1014						
QY	1074	CTAGCGAAAGGTCTTTTCGACGACGAGGAGGCTTACTTCTCGAAGCGGTGGCC	1127						
Db	1015	CTGTAAAGAGATCAACACGCTGACGGGACATCTGTACGAAGGCGGTGGCG	1068						

RESULT	6
V52155/c	
ID	V52155 standard; DNA: 28171 BP.
AC	V52155;
AD	V52155;
DT	23-OCT-1998 (first entry)
DE	Streptococcus pneumoniae genome fragment SEQ ID NO:22.
KW	Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KM	computer readable medium; vaccine; pharmaceutical composition; ds.
PN	Streptococcus pneumoniae.
PN	MO918931-A2.
PD	07-MAY-1998.
PF	30-OCT-1997; U19588.
PR	31-OCT-1996; US-029960.
PA	(HUMA-) HUMANA GENOME SCI INC.
PI	Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI	Kunsch CA, Rosen CA;
DR	WPI: 98-272225/24.
PT	Computer-readable medium with recorded Streptococcus pneumoniae
PT	polynucleotide sequences - useful in diagnostic kits and assays, and
PT	pharmaceutical compositions and vaccines for Streptococcus
PS	pneumoniae
PS	Claim 1: Page 273-289; 1409pp; English.
CC	The present invention describes a computer readable medium which has
CC	the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC	on it, or a representative fragment or a sequence at least 95% identical
CC	to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC	to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC	pneumoniae. The present invention also describes an isolated nucleic acid
CC	molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC	genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC	by a process comprising: (a) screening a genomic DNA library using as a
CC	probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC	to 391, identifying members of the library which contain sequences
CC	that hybridize to the target sequence and isolating the nucleic acid
CC	molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC	from an organism, amplifying nucleic acid molecules whose nucleotide
CC	sequence is homologous to amplification primers derived from the
CC	fragment of the S. pneumoniae genome to prime the amplification and
CC	isolating the amplified sequences. The computer readable medium can be
CC	used in a computer-based system for identifying fragments of the
CC	S. pneumoniae genome of commercial importance, or expression modulating
CC	fragments of the S. pneumoniae genome. Products from the present
CC	invention can be used in diagnosis kits and assays, and pharmaceutical
CC	compositions and vaccines for S. pneumoniae.
SQ	Sequence 28171 BP: 8507 A: 5080 C: 6440 G: 8144 T:

Query Match	16.5%;	Score 203.6;	DB 1;	Length 28171;
Best Local Similarity	54.0%;	Pred. No. 5.1e-34;		
Matches 534;	Conservative	0;	Mismatches 434;	Indels 20;
			Gaps	5;

QY	61	ATGCCGCTGATTTTCGACCGAGACCAAAACACGAATTCGGGTGGCCATCACCCG	120
Db	21711	ATGTTAATCGGAATCCAAAGAAATTAATAATACGAACCGTGTGCGCTTACACCT	21652
QY	121	GGCGGCTGGGGAATCAACCGCTGCTGGCATGAGTGTCTCATCGAGCGATCGGA	180
Db	21651	GCAAGTCTTACATCTAGTAAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	21592
QY	181	GAGGCTGCGCTATCACCGACGCGGATTTCAAGCGCGGCGGCAACGTCGGGAC	240
Db	21591	CTGCGTCTGCGCTTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	21532
QY	241	GCGACCAAGTGTGGGCGGACGCTGATTTATTCGTAAGGTCAAGAAACGATAGCGCG	300
Db	21531	GCTGTGTAAGCTTGGGCG--AGCAGAGTGTGTGTGAAATGAAGAACTTAAAGTCT	21475
QY	301	GAATACGCGCGCTGGCGACACGGGCG--GATCTGTGCTACGCTTTCGATTTGGCGGCTCA	359
Db	21474	GAATACGCTTACTTGGCGGACGATCTTCTTCTTCACTTTCGACACGAGCGCGCTGCT	21415

RESULT 6  
V52155/c  
ID V52155 standard; DNA: 28171 BP.  
AC V52155;  
DT 23-OCT-1998 (first entry)  
DE Streptococcus pneumoniae genome fragment SEQ ID NO:22.  
KW Streptococcus pneumoniae: S. pneumoniae: genome; diagnosis; assay;  
OS computer readable medium; vaccine; pharmaceutical composition; ds.  
PN Streptococcus pneumoniae.  
PD W09818931-A2.  
PF 07-MAY-1998.  
PR 30-OCT-1997; U19588.  
PA 31-OCT-1996; US-029960.  
PI (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,  
PI Kunesh CA, Rosen CA.  
DR WPI: 98-27225/24.  
PT Computer-readable medium with recorded Streptococcus pneumoniae  
PT polynucleotide sequences - useful in diagnostic kits and assays, and  
PT pharmaceutical compositions and vaccines for Streptococcus  
PT pneumoniae  
PS Claim 1: Page 273-289, 1409pp: English.  
CC The present invention describes a computer readable medium which has  
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
CC on it, or a representative fragment or a sequence at least 95% identical  
CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
CC pneumoniae. The present invention also describes an isolated nucleic acid  
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae  
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
CC by a process comprising: (a) screening a genomic DNA library using as a  
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
CC to 391, identifying members of the library which contain sequences  
CC that hybridize to the target sequence and isolating the nucleic acid  
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
CC from an organism, amplifying nucleic acid molecules whose nucleotide  
CC sequence is homologous to amplification primers derived from the  
CC fragment of the S. pneumoniae genome to prime the amplification and  
CC isolating the amplified sequences. The computer readable medium can be  
CC used in a computer-based system for identifying fragments of the  
CC S. pneumoniae genome of commercial importance, or expression modulating  
CC fragments of the S. pneumoniae genome. Products from the present  
CC invention can be used in diagnosis kits and assays, and pharmaceutical  
CC compositions and vaccines for S. pneumoniae.  
SQ Sequence 28171 BP; 8507 A; 5080 C; 5440 G; 8144 T;

QY 360 CGTGTGCAACCGATCGTTGGATTCCGCGACCACTGATTCCTACGAGACCGTC 419  
 DB 21414 CCAGATATAGACATCTATGTT-----AACAGCAAAAMCACTGAACTGTT 21367  
 QY 420 CAGACCGCCGAGGCGCATACCCCTGCTTCCCGCATAGGCAAGTCCCGGTGACAC 479  
 DB 21366 CGTGACATTCAGACACTACCGCTCTGTTCTATAGTAGAGTTCAGAGTGTATG 21307  
 QY 480 GCGCCGAGGTTGGCTTACCACTGATGCGAACCAGAGGCGCGCGGTGCTGATG 539  
 DB 21306 GCGTTC/ATGAGACTACTTCTTACTAGCAAGCTGGTGGCTGTGTTTACTT 21247  
 QY 540 GCGGGGTGCGCGCTCGAACCGCGGACGTCGTGTATGCGCGCGCGCGCGG 599  
 DB 21246 GGTGTGTACAGAGTTCCTCAAAAAGGAAAGTATATATGCTGTGTGCTGCT 21187  
 QY 600 TACAGACGAGCGCGCAACCGCAACGCGGCGCGCGCTTACGTTAGACATCAAC 659  
 DB 21186 ACAGATGTCGCGCGCATCCCTGCTGTGGTCTCAAGTACATTTTATATAGT 21127  
 QY 660 ATGACAACTTGGCACTCGACGCGGAGTCTGCGCGCG ATCCACACTGCTACTCA 719  
 DB 21126 TCCAACTGCTCTCACTTCTAGAGAGTCTTGGAACTCAATTCATCTATGCT 21067  
 QY 720 TCGGCTACGAGTGGAGGTGCGCAACGCGGCGGCTGTGATGCGCGCGCTGCT 779  
 DB 21066 ATTCATTTACATTTAGCAAGTGTGAGAGATGCTGATGTGTATGAGACATTC 21007  
 QY 780 GTGCCAGGCGCAAGGACCCAAATTTAGTCTGAAATTCCTTGTGCGGATGAAACA 839  
 DB 21006 ATCCCTGGTGAAGAACACCGAATGTGACAGATGATGATCAACAAATGCTGCA 20947  
 QY 840 GGTGGGCTGCTGTGATATGACATCGACGCGGCGGCTGTGCA---AGGCTACGA 896  
 DB 20946 GGTCTGTA-TGTTTACCTTCTGTTGACCAAGGCGGTATGGAACACCTGACCT 20888  
 QY 897 CCGACCACTACGACACCGGAGTTCGCGCTGACGACGCGCTTTTACTGCTGCG 956  
 DB 20887 GTGACAAACGAGATGAAACCGCTATATGAAACAGGGTGTCTCCATATGCGTTCC 20828  
 QY 957 AACATGCGCGCTGTGCGGAGAGCTGACCTGCGGCTTACCAAGCGACGATGCG 1016  
 DB 20827 AATATCCCTGCGGCTGCTGCTGCACTTCAACATGCGCTTACCAATGTCTTCT 20768  
 QY 1017 TATGTCTGAGCTTCCGACCATGCT 1044  
 DB 20767 TATATGAGAGCTTGGCTGCAAAAGAT 20740

RESULT 7  
 V74442/c  
 ID V74442 standard; DNA: 9280 BP.  
 AC V74442:  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SEQ ID #131.  
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 skin infection; surgical wound infection; scalded skin syndrome;  
 toxic shock syndrome; ds.  
 KW Staphylococcus aureus.  
 OS Key location/Qualifiers  
 PH misc-feature 841..900  
 FT /tag- a  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature 2641..2700  
 FT /tag- b  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT misc-feature 4441..4500  
 FT /tag- c  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature 6241..6300  
 FT /tag- d  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature 8041..8100  
 FT /tag- e  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

EP-786519-A2.  
 PD 30-JUL-1997. 100117.  
 PR 07-JAN-1997: 100117.  
 PR 05-JAN-1996: US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
 PI Rosen CA.  
 DR WPI: 97-374922/35.  
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 PT stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines  
 PS Claim 1: Page 710-715: 3271pp; English.  
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.  
 SQ Sequence 9280 BP: 2958 A: 1649 C: 1309 G: 3060 T:

Query Match 13.5%; Score 166.4; DB 1: Length 9280;  
 Best Local Similarity 50.5%; Pred. No. 2,4e-26;  
 Matches 483; Conservative 0; Mismatches 466; Indels 7; Gaps 3;

QY 151 CATAGGTGCTCATCCAGGAGCGCGGAGAGCGCTCGGCTATCACGACGCGGATTTC 210  
 DB 9280 CATAGCTTTTATGGAAGAAATGCGGGTTCAGGATCATCTTGAAGATGTGATTAC 9221  
 QY 211 AAGCGCAGGCGCGCACTGTGCGACCGGACGAGGTGGCGCGCGCTGATTATTA 270  
 DB 9220 AAGAAAGCAGGTGCTGATGATGTTGCTGAACAGCAAAAGTTGG---GATGGGATAG 9164  
 QY 271 TTGCTCAAGTCAAAAGACCGATAGCGCGGATACGGCGCTGCGACAGCGGC-GATC 329  
 DB 9163 GTTATTTAAAGTTAAAGACCACTGATTCGAATATTCATATTTTAAAGAGGCTTGA 9104  
 QY 330 TTGTCACGTTCTTGCATTTGGCGCGCTACGCTGTGACGCGATCGTTGATTTC 389  
 DB 9103 TTATTCCTTATCTTATTTAGCAATTAAGAAATTAACACAACTTGTGATGATAGA 9044  
 QY 390 GGCACCACTGCAATTCCTACAGAGCGTCCAGACCGGAAAGCGCACATACCTGCTT 449

```

Db 9043 AAGATATTAGTATTCATATATAGACTGTGCACTACCAAGCCGACTTTTACATTGTTA 8984
QY 450 GCCCGATGAGCGAAATGCGCGGTGACTGCGCCGAGGTTGGCGCTTACCACTGATG 509
Db 8983 TCACCAATGAGTAGGATAGCAGAGAAATGTGAGCTCAAGTTGGCGAGATTCCTACAA 8924
QY 510 GCAACCCAGGGGGCGCGGTGTGATAGGGCGGGTCCCGGGGTGCAACCGCCGAC 569
Db 8923 AACTTAAATGAGTATGAGGAAATTTACTAGGTGTGTCCGAGAGTACTTAAAGGTTAA 8864
QY 570 GTGTGTGATGAGCGCGCCGCGACCCGCGCTACAAAGCAGCCGCGGATGCGCAAGCATG 629
Db 8863 GAACTATATATGAGTGTGCTGCTCAAGCAAGAAATGCGCTAAATTTGCACTAGACTA 8804
QY 630 GCGCGAGCGCTTACGCTTATACATCAACATGCAACAACTTGGCAACTGCAAGCCGAG 689
Db 8803 GGTGAGATGATTCATTTTAAATTTATCCMAAGCGTTTCAACAACTTAACTATTTA 8744
QY 630 TTCTGCGCGCGATCCACACTGCTACTCATGCGCTTACAGCTGAG, GTGCGGTCAA 749
Db 8743 TTCGGTGGAGCTGTACATATATATGCAAAATCCGTGAAATATGATGTGATTTAA 8684
QY 750 CGTCCGAGCTGATGATGAGGCGCTCCTGTGCGAGCGCGCAAGCAACCAATTAATC 809
Db 8683 CAAGTGTATTTAGTATATAGTGTCAAGTTTAAATCCAGGTGTAAGCGCAAGACTTGA 8624
QY 810 TCGAATTCATCTGTGCGCATATGAAACCAAGTGGGTAAGTGTGATATAGCTTCAAC 869
Db 8623 ACAGAGACATATTTAAACAAATGAAATGGGTGATGATATATGCACTTGC, ATGAT 8564
QY 870 CAGGCGGCGTGTTCGAAGGCTC---ACGACCGCACCTACAGCACCGCGAGCTTCCG 926
Db 8563 CAAGCGGATTTTGAACAACATGATAAATATAGACATGATGATGCTTACATATAT 8504
QY 927 GTGCAAGACACCTGTTTACTGCTGGGCAACATGCGCGCTCGGTGCGCAAGCTG 986
Db 8503 AAGCATGTGTGTATATGATGATGATGATGATGATGATGATGATGATGATGATG 8444
QY 987 ACCTACGCGCTGACCAACGCGAGTGGCGATGCTGTGCTGATGCTGCGAGCATGCGG 1046
Db 8443 ACGTATAGCTTTAAATATATGCTATGCTATGCTATGCTATGCTATGCTATGCTAT 8384
QY 1047 CCGGCGCGCTGCGCTGCAATCGGCACTAGGCAAAAGCTTTTCGACAGCAAG 1102
Db 8383 AGAGAGCATTTAAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 8328

RESULT 8
T67971 standard: DNA: 1074 BP.
AC T67971;
DT 15-JUL-1997 (first entry)
DE H. pylori membrane protein ORF 05CP20518orf1.
KW vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW membrane; amino acid; metabolism; ds.
OS Helicobacter pylori.
FH key Location/Qualifiers
FT cds 1..1074
FT /note="no stop codon given"
PN W09640893-A1.
PD 19-DEC-1996.
PR 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR) ASTRA AB.
PI Berglund OT Smith D, Mellgaard BL;
DR WPI: 97-052306/05.
PT P-SDS; W20718.
Helicobacter pylori nucleic acid sequences and related
polypeptide(s) - useful for vaccines to treat or prevent H. pylori

```

```

PT infection, and to detect Helicobacter
PS Claim 27; Page 819; 1481p; English.
CC The present sequence encodes a Helicobacter pylori membrane
CC protein likely to contain four membrane spanning regions.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the
CC bacterial DNA. The sequences were analysed for ORF of at least 180
CC nucleotides, and the predicted coding regions defined by computer
CC evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or expected
CC membrane proteins. Having identified and determined the sequences
CC of interest, particular regions can be isolated from H. pylori by
CC PCR amplification for recombinant polypeptide production, e.g. in
CC E. coli hosts.
SQ Sequence 1074 BP: 302 A; 189 C; 297 G; 286 T;

```

```

Query Match 7.2%; Score 89.2; DB 1; Length 1074;
Best Local Similarity 46.5%; Pred. No. 2e-10;
Matches 373; Conservative 0; Mismatches 413; Indels 16; Gaps 2;

```

```

QY 265 GATTATTTGCTCAAGCTCAAGAACCAATGAGCGGCAATACGCGCTGCGA-CAGG 333
Db 133 GATTGTGGTCAAAATGCAAAAGACCTTTAGCATGATACCTTTGCTCAAGAAAA 192
QY 324 GCGATCTGTGCAAGTCTGTGCAATTTGGCGCGTACAGTGTGCTACCGATGGTGTG 383
Db 193 GCGACTGCTTATATTTATTTGATTTAGCTATCAAAAAAGCTGTGTGAATGTTATT 252
QY 384 GATTCGCGACACAGCTCAATTCCTACAGAGACGCTCCAGACCGCGCAAGCGCACTACC 443
Db 253 AATAAAAAATCACTCTTATTTGCACTGAACCATTCGCGGCGCTAAAGACATACCT 312
QY 444 CTGCTTGGCGGATGACGAGTGGCGGCTGACTGCGCGCGCCAGAGTTGCC----- 495
Db 313 ATTTTACGCGCTTATGACGCTGTGGAGAGTTGCTGCTGCTGCTGCTGCTGCTGCT 372
QY 495 -----GCTTACCACTGATGCGAACCAAGGGGCGCGGTGTGCTGATGGGGGGTG 548
Db 373 TTACTGCTTTAGAGCATGTTAAAGGTTATGAGGTAAAGGGGCTCATGCTAAGGGGTTG 432
QY 549 CCGGCGCTGCAACCGCGCAAGCTGTGATCGCGCGCGACCGCGGCTCAACAGCA 608
Db 433 TCGGTCGCGCAAGGCTTAATATGCTGTAATGAGAGCGGTGCTGCTGCTGCTGCTG 492
QY 609 GCGCGATCGCAAGGCTGAGCGGCGGCGGCTTACGCTTACAGCATCAATGACAA 668
Db 493 GCGAAAGTCTTAAAGCAAAATGGGGCTTAAAGTAAAGATTTTGAATTAAGCTAGCTAA 552
QY 669 CTTGCGCACTGACGCGGAGTGTGCGCGGATCCACACTGCTGCTATAGGCTTAC 728
Db 553 TTCAAAAACCACTTATATATTTATGATTTAAGATTTAAGCTTAAAGGATGAAGCC 612
QY 729 GAGCTCAAGGTGCGGCAACGTCGACCTGTGATTTGGGCGCTCTGTGCGAGC 788
Db 613 AATATCAATCAAGCTTAAACGGGCGGTGAGGCTATGAGACGCGGTGCTTACAGG 672
QY 789 GCCAAGCGACCAATATGCTGCAATTCAGTGTCCGCAATGAAACAGGCTGGCGTA 848
Db 673 AGCAAAACCCCTTAAGTATCTTAAGAAAGCAATTTAAATATACATGACAGACAAAGGGTA 732
QY 849 CTGTGATATATGCAATGACGACGAGGCGCTGTTGCAAGGCTCAGACGACCACTTAC 908
Db 733 GTATTTATGTGCTTGCAGTTTAAAGGGGCTGCATTAAGACCATACGCAACCAAT 792
QY 909 GACCACCGAGCTGTGCGCGTGCAGACGACGCTTTACTGCTGCGGCAACATGCGCGCG 968
Db 793 TCTAACCGGTGTATGTGAGAAAGATTTTCTTCATTTATGCGGTGCGGACATGCGCAGG 852

```







FT misc-feature 19201. 19260  
 FT /\*tag- k  
 FT /note- "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence".  
 FT misc-feature 21001. 21060  
 FT /\*tag- l  
 FT /note- "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence".  
 FT misc-feature 22801. 22860  
 FT /\*tag- m  
 FT /note- "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence".  
 FT misc-feature 24601. 24660  
 FT /\*tag- n  
 FT /note- "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence".  
 FT misc-feature 26401. 26460  
 FT /\*tag- o  
 FT /note- "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence".  
 FT misc-feature 28201. 28260  
 FT /\*tag- p  
 FT /note- "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence".  
 FT misc-feature 30001. 30060  
 FT /\*tag- q  
 FT /note- "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence".  
 EP-786519-A2.  
 PD 30-JUL-1997.  
 PF 07-JAN-1997; 100117.  
 PR 05-JAN-1996; US-009861.  
 PA (HUMAN-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunach CA,  
 PI Rosen CA;  
 DR WPI: 97-374922/35.  
 PT Polynucleotide(s) and proteins derived from *Staphylococcus aureus*  
 PT stored on computer readable medium and used in the production of  
 PT anti-S. aureus vaccines  
 PS Claim 1; Page 452-469; 3271pp; English.  
 CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S. aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S. aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S. aureus in a sample. S. aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S. aureus DNA sequences contained on the  
 CC computer readable medium.  
 SQ Sequence 31096 BP; 11857 A; 5243 C; 5477 G; 7488 T;

Query Match 4.6%; Score 56.8; DB 1; Length 31096;  
 Best Local Similarity 48.2%; Pred. No. 0.0013;  
 Matches 160; Conservative 0; Mismatches 172; Indels 0; Gaps 0;  
 OY 713 CTACTCATCGGCTTACGAGCTCGAGGGTCCGTCMAACGTCGACCTGTGATGGGCG 772  
 DB 649 CAATCAACACCCAGGAAATTTAGCAGAACAAATTAAAGAGAGATGATTTATTTAC 708  
 OY 773 CGCTGTGGCAGGCGGCGGCAAGCCAAATTAATTCGATTCGATTCGCGGCAT 832  
 DB 709 AATTATTTACAGTGGGAAACCGCCAAATTTGTTACTGCTGAGATGTTAATCAAT 768  
 OY 833 GAACCCAGTGGCTACTGCTGATATATAGCATGACGAGGCGGCTTTGGAAGGCTC 892  
 DB 769 GAAAAAAGTTCAGATTAATTCGATATATAGCATATGACAGGTGATTCATTTAGTGT 828  
 OY 893 AGACCCGACACCTACGACACCCGACGCTGCGGTGACGACACGCTGTTTACTGCGT 952  
 DB 829 TAGACCAACTACAAATTTCTGATCCAGTGTATGAAAGAGAGGTGATTCATTTAGTGT 888  
 OY 953 GGGCAACATGCCCGCTGCGTCCGCAAGACGTCGACCTACGCGGTGACCAAGGACGAT 1012  
 DB 889 ACCAATATCAACGAGGAGGAGTCCCAAGACCTTCANCAATGGCATTTAGCAAGAAATAT 948  
 OY 1013 GCCGATGTGCTGACGCTTGCCGACCATGCT 1044  
 DB 949 TGATTATATTATAGAAATTTGTGACAAAGGCT 980  
 RESULT 11  
 Q21833/C  
 ID Q21833 standard; DNA: 390 BP.  
 AC Q21833;  
 DT 08-JUN-1992 (first entry)  
 DE Randomising oligonucleotide used in SPERT mRNA prepn.  
 KW Systematic polypeptide e volution by reverse translation; SPERT.  
 KW ligand binding; ss.  
 OS Synthetic.  
 PN W09202536-A.  
 PD 20-FEB-1992.  
 PF 01-AUG-1991; U05463.  
 PR 02-AUG-1990; US-561968.  
 PA (COLS.) UNIV OF COLORADO.  
 PI Gold L, Thert C;  
 DR WPI: 92-080018/10.  
 PT New method of systematic polypeptide evolution by reverse  
 PT translation - by linking each polypeptide in sample mix. to  
 PT individualised mRNA allowing further synthesis of selected  
 PT polypeptide(s)  
 PS Example; Page 55; 102pp; English.  
 CC The sequence is that of an example randomising oligonucleotide which  
 CC is used in the prepn. of mRNA encoding candidate polypeptides for the  
 CC method of systematic polypeptide evolution by reverse translation  
 CC (SPERT). The method provides a rapid way of isolating and identifying  
 CC polypeptide ligands which bind to target mols. The polypeptide ligands  
 CC can be used in e.g. assay methods, diagnostic procedures, cell sorting,  
 CC as activators or inhibitors of enzyme act., function, as probes, as  
 CC sequencing agents, drug delivery vehicles, modifiers of hormone  
 CC action and as catalysts. see also Q21830-Q21832.  
 SQ Sequence 390 BP; 125 A; 126 C; 133 G; 6 T;  
 Query Match 4.0%; Score 49.2; DB 1; Length 390;  
 Best Local Similarity 45.9%; Pred. No. 0.036;  
 Matches 168; Conservative 0; Mismatches 198; Indels 0; Gaps 0;  
 OY 425 CGCGAAGGCGCACTACCCCTGCTGCCCCGATGAGCAAGTCGCGTCACTGCGCGC 484  
 DB 366 CGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 307  
 OY 485 CCAGTGTGGCGCTTACACACCTGATGCAACCAAGGGGCGCGGTGCTGATGGCGG 544





Fri Jun 23 09:31:38 2000

us-09-362-485-8.rng

Page 12

Search completed: June 22, 2000, 15:10:52  
Job time: 17978 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:13:36 ; Search time 341.15 Seconds

(without alignments)  
470.560 Million cell updates/sec

Title: US-09-362-485-8

Perfect score: 1335

Sequence: 1 ATCTGAGATGATCGAAC.....GGAGGCGATGATCGGCC 1235

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/1na/5C.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/1na/5D.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/1na/6.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/1na/PCRM.COMB.seq.\*  
7: /cgn2\_6/ptodata/2/1na/Backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45.4	3.7	833	US-08-403-852D-3	Sequence 3, Appl
2	45.4	3.7	5392	US-08-403-852D-1	Sequence 1, Appl
3	44.4	3.6	2582	US-08-816-105A-2	Sequence 2, Appl
4	43.6	3.5	28958	US-08-258-261B-6	Sequence 6, Appl
5	43.6	3.5	28958	US-08-456-837-6	Sequence 6, Appl
6	43.6	3.5	28958	US-08-457-342-6	Sequence 6, Appl
7	43.6	3.5	28958	US-08-457-646A-6	Sequence 6, Appl
8	43.6	3.5	28958	US-08-458-076A-6	Sequence 6, Appl
9	43.6	3.5	28958	US-08-764-233A-4	Sequence 6, Appl
10	43.6	3.5	28958	US-08-457-335A-6	Sequence 6, Appl
11	43.6	3.5	28958	US-08-729-214-6	Sequence 6, Appl
12	43.6	3.5	49377	US-08-764-233A-1	Sequence 6, Appl
13	42.4	3.4	1833	US-08-403-852D-6	Sequence 6, Appl
14	42.4	3.4	2185	US-08-173-508-3	Sequence 3, Appl
15	42.4	3.4	2185	US-08-265-310-3	Sequence 3, Appl
16	41.6	3.4	1620	US-08-461-775-10	Sequence 10, Appl
17	41.6	3.3	20235	US-07-642-734C-3	Sequence 3, Appl
18	40.6	3.3	4257	US-08-690-473-1	Sequence 1, Appl
19	40.6	3.3	12001	US-08-458-568A-11	Sequence 11, Appl
20	39.6	3.2	2414	5248599-1	Patent No. 5248599
21	39.6	3.2	2668	US-08-461-775-11	Sequence 11, Appl
22	39.4	3.2	459	US-08-387-942C-35	Sequence 35, Appl
23	39.4	3.2	1978	5212296-8	Patent No. 5212296
24	39.4	3.2	12588	US-08-387-942C-1	Sequence 1, Appl
25	39.2	3.2	420	US-08-470-175-148	Sequence 148, App
26	39.2	3.2	44377	US-08-804-227C-7	Sequence 7, Appl
27	39.2	3.2	44377	US-08-804-198-1	Sequence 1, Appl

28	39	3.2	43280	3	US-08-804-227C-1	Sequence 1, Appl
29	38.8	3.1	8051	4	US-08-576-626A-2	Sequence 2, Appl
30	38.8	3.1	11219	2	US-07-642-734C-1	Sequence 1, Appl
31	38.6	3.1	2064	1	US-08-343-428-1	Sequence 1, Appl
32	38.6	3.1	11219	2	US-07-642-734C-1	Sequence 1, Appl
33	38.4	3.1	1215	4	US-08-947-726A-1	Sequence 1, Appl
34	38.4	3.1	2048	5	US-08-776-251-1	Sequence 1, Appl
35	38.2	3.1	2109	4	US-08-555-568B-20	Sequence 20, Appl
36	38	3.1	474	3	US-08-403-852D-14	Sequence 14, Appl
37	38	3.1	30001	3	US-08-125-468-1	Sequence 1, Appl
38	38	3.1	30001	3	US-08-474-933-1	Sequence 1, Appl
39	37.8	3.1	3252	4	US-08-809-740A-1	Sequence 4, Appl
40	37.8	3.1	3252	4	US-08-809-740A-4	Sequence 4, Appl
41	37.8	3.1	15664	1	US-08-402-282-3	Sequence 3, Appl
42	37.8	3.1	15664	1	US-08-508-004-3	Sequence 3, Appl
43	37.8	3.1	15664	1	US-08-402-068-3	Sequence 3, Appl
44	37.8	3.1	15664	1	US-08-402-068-3	Sequence 3, Appl
45	37.6	3.0	1524	7	5512669-1	Patent No. 5512669

## ALIGNMENTS

RESULT 1  
US-08-403-852D-3  
Sequence 3, Application US/08403852D  
Patent No. 5891695  
GENERAL INFORMATION:  
APPLICANT: Blanc, Veronique  
APPLICANT: Blanche, Francis  
APPLICANT: Crouzet, Joel  
APPLICANT: Jacques, Nathalie  
APPLICANT: Lacroix, Patricia  
APPLICANT: Thibaut, Denis  
APPLICANT: Zagorec, Monique  
APPLICANT: Debussche, Laurent  
TITLE OF INVENTION: De Crecy-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved In The  
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,852D  
FILING DATE: 10-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806.0054-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 833 base pairs  
TYPE: nucleic acid









COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,342  
FILING DATE: 01-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/457,205  
FILING DATE: 01-JUN-1995  
APPLICATION NUMBER: 08/258,261  
FILING DATE: 08-Jun-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28958 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
US-08-457-342-6

Query Match 3.5%; Score 43.6; DB 1; Length 28958;  
Best Local Similarity 44.8%; Pred. No. 0.097;  
Matches 290; Conservative 0; Mismatches 354; Indels 4; Gaps 3;  
QY 148 GGCATGAGTGTCTTCAGGAGGCTGCGGAGAGGCTTTCACCGAGCGGAT 207  
DB 25306 GCCACGACGACCTACCTCCGGGATGTGGCTATTCCCTGGCCACACCGCGCCAC 25365  
QY 208 TTCAGAGGGGAGCGCCGACACGTCGCGACCGCCGACCAAGTGTGGCCGACGCTGAT 267  
DB 25366 TTGAGACACCGCGCCCTCTCGAGCCACACCGGAGAGGCTCTCTCCGCGCTCGAC 25425  
QY 268 TTATGCTCAAGTCAAA-AGAACCGATAGCGGCGGATAGCGCGCCCTCGACACGCGG 326  
DB 25426 TGGCTGGCCAGAGACAGCCCGCCGACACCGTCTCGGAGGAGCGGAAGCCACGCG 25485  
QY 327 ATCTTGTACGTTCTTTCGATTTGGCCGCTACGTCGTCGACCGATCGTTGTTGAT 386  
DB 25486 AAGCTGCTTGTCTTCTTCTGCGCAAGGCTGCAAGTGGAAAGGATGCGCTCTGCTG 25545  
QY 387 TCCGACACGATGCAATTCCTACGACCGGTCACCGCCGAGGAGGAGGAGGAGGAGG 446  
DB 25546 CTCGACTCTCTGCGCCCTCTTCGCGGCTCAGCTGGAAGCATGCGAGCGCGCTCGCT 25605  
QY 447 CTGGCCCGATAGGAGGAGTCCCGGCTGCTGCGCCCGCCAGTTGGCCTTACGAC--C 504  
DB 25606 CACGTCGAGTGGAGCGCTGCTGCGGCTGCGCGCGGAGGAGGCGCGCCCTCCGAC 25665  
QY 505 TGAATCGAACCAAGGAGGCGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 564  
DB 25666 CCGCTGAGCTGACAGCGCGCCCTCTTTCGCGTATGCTCTCCCTGCGGCGCTCTG 25725  
QY 565 CCGAGC-TGCTGATGATCGCGCGCGGACCGCGGCTACACGACCGCGATCGCAAC 623  
DB 25726 CGCTGCGTGGGAGGAGCGCGCGCGGCTGCGGACAGTACAGGAGGAGGAGTCCGCG 25785  
QY 624 GGCATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 683  
DB 25786 GCGTGTGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 25845  
QY 684 GCGGAGTCTGCGGCGGAGTTCACACTGCTACTATGAGGCTAGAGGCTGAGGAGTGC 743

DB 25846 AAGCGCTCACCACCGTCCCGGCAACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 25905  
QY 744 GTCAAACGTCGAGCTGTGATTTGGGCGGCTTCGTCGACGAGCGCC 791  
DB 25906 GACCTTCAGACCTTACCTCCCTCCCTGGGCGAGAGGCTCTCATCGCC 25953

RESULT 7  
US-08-457-646A-6  
Sequence 6, Application US/08457646A  
Patent No. 5676550  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James M.  
APPLICANT: Beck, James Joseph  
APPLICANT: Hill, Dwight Steven  
APPLICANT: Ryals, John Andrew  
APPLICANT: Gaffney, Thomas Deane  
APPLICANT: Lam, Stephen Ting  
APPLICANT: Hammer, Phillip E.  
APPLICANT: Uknes, Scott Joseph  
TITLE OF INVENTION: Genes for the synthesis of  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,646A  
FILING DATE: 01-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/457,205  
FILING DATE: 01-JUN-1995  
APPLICATION NUMBER: 08/258,261  
FILING DATE: 08-Jun-1994

ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28958 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
US-08-457-646A-6

Query Match 3.5%; Score 43.6; DB 1; Length 28958;  
Best Local Similarity 44.8%; Pred. No. 0.097;  
Matches 290; Conservative 0; Mismatches 354; Indels 4; Gaps 3;

QY 148 GGCATGAGTGTCTTCAGGAGGCTGCGGAGAGGCTTTCACCGAGCGGAT 207  
DB 25306 GCCACGACGACCTACCTCCGGGATGTGGCTATTCCCTGGCCACACCGCGCCAC 25365  
QY 208 TTCAGAGGGGAGCGCCGACACGTCGCGACCGCCGACCAAGTGTGGCCGACGCTGAT 267

```

Db 25366 TTCGAGCAGCCGCGGCTCTGTGTAACCAACCGCGAGCTCTCTCCGCGCTGAC 25425
Qy 268 TTATGCTCAAGGTCA-A-AGAAGCATAGCGCGGAATACGCGCCCTCGACACAGGGCG 326
Db 25426 TCGCTCGGCCAGAGAACCGCGCCCGACCGCTCTCGAGAGAGGAGGAAAGCAGCGC 25485
Qy 327 ATCTGTTCACGTTCTTGTGATTTGGCCCGCTGACGCTGTTGACCGCATGCTTGTGGAT 386
Db 25486 AAGCTGCTTCTGCTTCTCTGAGGAAAGCTGCGAGTGGGAAAGGAGTCCCTCTGCTG 25545
Qy 387 TCCGACACACGATCAATTGCTGACAGACCGCTCAGACCGCGCGACACTACCTG 446
Db 25546 CTCGACTCTCGCCCGCTCTCCGCGCTCAGCTCGAAGATGAGAGCGCGCGCTGCTCT 25605
Qy 447 CTGCGCCGATGAGGAAAGTCGCGGCTGACTCGCCCGCCAGATTGCGCTTACAC--C 504
Db 25606 CACGTCGAGTGAAGCTCTGCTCTCTGCGCGCGAGAGGCGCGCCCTCTCTGAC 25665
Qy 505 TGATCGGAACCCAAAGGGGCGGCTGCTGATGGGCGGGGTCGCGCGATCGAACCG 564
Db 25666 CGCTGAGAGTGTGACAGCCGCTCTTTCGCTCATGCTCTCTGCGCGCTCTG 25725
Qy 565 CCGAGC-TGCTGTGATGCGCGCGCGACCGCGGCTCAACGACGCGCATCGCCAAC 623
Db 25726 CGCTGCTCGGCTGAGAGCGCGCGCGCTGCGCGACAGTCAAGGAGGATCGCGCC 25785
Qy 624 GGCATGCGCGCGCGCGCTTACGTTTACATCAACATCGACAACTTCGCGCACTCGAC 683
Db 25786 GCGTCTGCTGCGAGGGGCTCTCTCTCGAGAGCGGGCGCGCTGCGCGCTGCGAGC 25845
Qy 684 GCGGAGTCTGCGCGCGCGCTCTCTCTCGAGAGCGGGCGCGCTGCGCGCTGCGAGC 743
Db 25846 AAAGCGCTCACACCGCTGCGCGCGCAAGCGGGCGAGCGCGCGCTGCGCGCTGCG 25905
Qy 744 GTCAACGTCGCGCGCTGCTGATGGGCGCGCTCTGCTGCGCGCGCTGCGCGC 791
Db 25906 GACCTCGACGACTACCTGCTCTCTGCGCGCGAGAGGCTCTCATCGCC 25953

```

```

RESULT 8
US-08-458-076A-6
: Sequence 6, Application US/08458076A
: Patent No. 5698425
:
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Lyon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 St. Line Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458.076A
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 435

```

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-JUN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEetical: NO
: ANTI-SENSE: NO
:
: US-08-458-076A-6
:
: Query Match 3.5%; Score 43.6; DB 2; Length 28958;
: Best Local Similarity 44.8%; Pred. No. 0.097;
: Matches 290; Conservative 0; Mismatches 354; Indels 4; Gaps 3;
:
: Qy 148 GGCATGAGTGTCTATTCAGGAGGTGCGGAGAGGCTGCTATCAGCGAGCGAT 207
: Db 25306 GCCCAGCAGACCTCACCCCTGCGGATGTGGCTATTGCTGCGCACACCGCGCCAG 25365
: Qy 208 TTCAGC TGGAGGGGCGGCACTGTGTGACCGCGCGCGACCGAGTGTGGCGAGCTGAT 267
: Db 25366 TTGAGC TCGAGCGCGCTCTGCTGAGCCACACCGGAGAGCTCTCTGCGCTCGAC 25425
: Qy 268 TTATGCTCAAGGTCA-A-AGAAGCATAGCGCGGAATACGCGCCCTCGACACAGGGCG 326
: Db 25426 TCGCTCGGCCAGAGAACCGCGCCCGACCGCTCTCGAGAGAGGAGGAAAGCAGCGC 25485
: Qy 327 ATCTGTTCACGTTCTTGTGATTTGGCCCGCTGACGCTGTTGACCGCATGCTTGTGGAT 386
: Db 25486 AAGCTGCTTCTGCTTCTCTGAGGAAAGCTGCGAGTGGGAAAGGATCCCTCTGCTG 25545
: Qy 387 TCCGACACACGATCAATTGCTGACAGCGCTCAGACCGCGCGAGAGCGCATACCCCTG 446
: Db 25546 CTCGACTCTGCGCGCTCTCTCTGCGCTCAGCTCGAAGCATGAGCGCGCTGCTCTCT 25605
: Qy 447 CTGCGCCGATGAGGAAAGTCGCGGCTGACTCGCCCGCGAGTTGGCGCTTACAC--C 504
: Db 25606 CAGTCTGAGTGAAGCTGTGCTGCTCTGCGCGCGAGAGGCGCGCGCTGCTCTGAC 25665
: Qy 505 TGATCGGAACCCAAAGGGGCGCGGCTGTGATGGGCGGGTGGCGCGCTGGAACCG 564
: Db 25666 CGCTTCACGCTGACAGCGCGCGCTCTTTCGCTCATGCTCTCTGCGCGCGCTCTG 25725
: Qy 565 CCGAGC-TGCTGTGATGCGCGCGCGACCGCGGCTCAACGAGCGCGCATCGCCAAC 623
: Db 25726 CGCTGCTCGGCTGAGAGCGCGCGCGCTGCGCGACAGTCAAGGAGGATCGCGCC 25785
: Qy 624 GGCATGCGCGCGCGCGCTTACGTTTACATCAACATCGACAACTTCGCGCACTCGAC 683
: Db 25786 GCGTCTGCTGCGAGCGCTCTCTCTGAGAGAGCGCGCGCGCATGCGCGCGCTGAGC 25845
: Qy 684 GCGGAGTCTGCGCGCGCGATTCACACTGCTTACTCATGCGCTTACAGCTGAGGCTGC 743
: Db 25846 AAAGCGCTCACACCGCTGCGCGCGCAAGCGGGCGAGCGCGCGCTGCGCGCTGCG 25905
: Qy 744 GTCAACGTCGCGCGCTGCTGATGGGCGCGCTCTGCTGCGAGCGCGC 791
: Db 25906 GACCTCGACGACTACCTGCTCTCTGCGCGCGAGAGGCTCTCATCGCC 25953

```

```

RESULT 9
US-08-764-233A-4
: Sequence 4, Application US/08764233A
: Patent No. 5716849
: GENERAL INFORMATION:
: APPLICANT: Ligon, James M.
: APPLICANT: Schupp, Thomas
: APPLICANT: Beck, James J.
: APPLICANT: Hill, Dwight S.
: APPLICANT: Neif, Snezana
: APPLICANT: Rials, John A.
: TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 520 White Plains Road, P.O. Box 2005
: CITY: Tarrytown
: STATE: NY
: COUNTRY: USA
: ZIP: 10591
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/764,233A
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/729,214
: FILING DATE: 09-OCT-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/258,261
: FILING DATE: 08-JUN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Weigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: 1506/CIP6
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (919) 541-8587
: TELEFAX: (919) 541-8689
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Sorangium cellulosum
: IMMEDIATE SOURCE:
: CLONE: p98/1
US-08-764-233A-4

```

[illegible]

```

Db 25486 AAGTCGTCCTTTCTTTCTTCTTG6GCAAGGCTGCGAGTG6GAAGGATG6GCTTCGTGCGT 255435
QY 387 TCGGGACACAGCAATTGCTTACGAGACCGGTCCAGACC6CCGAAGGCGGCACTACCCCTG 446
Db 25546 CTCGACTTCCTCGCCCGCTTCCGCGCTAGCTCGAAGCATCGAAGCGCGGCGTGGCTCT 25605
QY 447 CTTGCCCCGATGAGCGAGTGGCCGCTGCACTCGCCGCCCAAGGTTGGCGCTTACAC--C 504
Db 25606 CACCTCGAGTGGAGCTGGTCTCCCGTCTCGGCCCCCGACGAG6GGGCGCCCCCTCCCTCGAC 25665
QY 505 TGATGCCAACCACAAGGGGGGCGCGGCTGTGTGATGAGGGGGGCTGCCGGGCTGGAACCG 564
Db 25666 CGCGTCGACGCTCTACAGCCGCCCTCTTTGCGGTCATGCTCTCCCTGGGCGGCTCTTG6 25723
QY 565 CCGAGC--TGTGTGTATCGGCGCGGCGGACCGCGGCTCAACAAGGACGCCGATCGGCAAC 623
Db 25726 CGCTCGGCTCGGGCTAAGAGCGCCGCGCGCGCTGTCGCGCACATGAGGGCGAGATCGCGCG 25785
QY 624 GCGATGGGCGGCGGACGCTTACGGTTTACGACATCAACAATCGACAACTTGGCAACTCGAC 683
Db 25786 GCGCTTCGTCGAGGCGCTCTCTCCCTCGAGGAGCGGCGCCCAATCGCGGCTCGCCACG 25845
QY 684 GCGGAGTTCTGCGCGCGGATCCACACTCGCTACTCATCGGCTACGAGCTCGAGGGTGGC 743
Db 25846 AAACGGCTCACACAGCTGCGCGGCAACGGGGCAGGCGGCGGCTCGAGCTCGGCGCTCC 25905
QY 744 GTCAACGTCGCGACCTGTGTATTGGGGCGCTCGTGTGTGCGAGGCGCC 791
Db 25906 GACCTCGAGACTACTCGTCCCTCGTGGGCGACAGGCTCTCATCGCC 25953

RESULT 10
US-08-457-335A-6
; Sequence 6, Application US/08457335A
; Patent No. 5723759
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,335A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,251
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129

```

REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
 TELEPHONE: 919-541-8614  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 28958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-457-335A-6

Query Match 3.5%; Score 43.6; DB 2; Length 28958;  
 Best Local Similarity 44.8%; Pred. No. 0.097;  
 Matches 290; Conservative 0; Mismatches 354; Indels 4; Gaps 3;

OY 148 GGCATGAGTGTCTCATCCAGCAGTGGCGGAGAGGGCTGCGTATACCGAGCGGGAT 207  
 DB 25306 GCGCCAGAGACCTTACCTCCGCGATGCGCTATTGCTGTGGCAACCGCGCCGAC 25365  
 OY 208 TTCAAGGGGCGAGCGCGCAACTGTGTGGCACCGGACCAAGTGTGGCGCGACCTGAT 267  
 DB 25366 TTGAGAACCGGGCGGCTCTGTAGCCCAACCGGAGAGAGCTCTCTCGCGGTGAC 25425  
 OY 268 TTATTGCTCAAGTCAA-AGAACCATAGGGCGGAAATACGGCCGCTTCCGACAGGGCG 326  
 DB 25426 TCGCTCGCCGACGAGAACCGCCCGGACACCGTCTCGAGGAGGAGGAGCCACGAC 25485  
 OY 327 ATCTGTTCACGTTCTTGCATTTGGCCGCGGCTGACGTGCTTGCACGATGCTGTGGAT 386  
 DB 25486 AAGCTCGTCTTCTGCTTCTTCTGCGCAAGGCTGCGATGGGAAGGATGGCCCTTCCG 25545  
 OY 387 TCCGGCACCACTCAATGCTCTAGAGACCGTCCAGACCGCGGAAGGCGACTACCCCTG 446  
 DB 25546 CTCGACTCTCTGCGCGCTTCCGCTCAGTCAAGATGCAAGGAGGCGGCGCTCTCT 25605  
 OY 447 CTTGGCCCGATGAGGAGTCCCGGTGACTGCTGCGCCCGCCAGTTGGGCTTACGAC--C 504  
 DB 25606 CACGTCGAGTGAAGCTCTGCTCCGCTCGCCGCGGAGAGGGGCGCCCTCTCTGAC 25665  
 OY 505 TGAATCGAAGCCCAAGGGGCGCGGTGTGCTGATGAGCGGGGTGCGCGCGTCAAGCG 564  
 DB 25666 CCGCTCGAGCTGTACAGCCCGCCCTTGTGCGGTATGTCTCTTGGCGGCGCTTGG 25725  
 OY 565 CCGACG-TGCTGTGATGCGCGCGGCGACCGCGGCTACAGCAGCGCCGATGGCAAC 623  
 DB 25726 CGCTGCTGCGGTAGAGCCCGCGCGCTGCGGACAGTCAAGGGGCGATCGCGCC 25785  
 OY 624 GGCATGGGGGCGAGGTAAGTCTTAGACATCAACATGCAAACTTGGGAACCTGAC 683  
 DB 25786 GCTTCTGTGAGGCGGCTCTCTCTCGAGAGCGGGCCGCGATGCGCGCTTGGAC 25845  
 OY 684 GCGGATTTGGCGCGGATCCACATCTGCTATCATGCGCTAGAGAGTCAAGTGGC 743  
 DB 25846 AAGGCGCTACACACGCTGCGCGGCAAGGGGCGCATGGCGCGCTGTGAGAGTCTG 25905  
 OY 744 GTCAAAAGTGGCGAGCTGTGATTGGGGCGGCTGTGTC 366GCC 791  
 DB 25906 GACCTCCAGACCTACCTGCTCTCTGGGCGGACAGGCTCTCATCGCC 25953

RESULT 11  
 US-08-729-214-6  
 Sequence 6, Application US/08729214  
 Patent No. 5817502  
 GENERAL INFORMATION:  
 APPLICANT: Ligon, James M.  
 APPLICANT: Hill, Dwight Steven  
 APPLICANT: Ryals, John Andrew

APPLICANT: Hammer, Phillip E.  
 APPLICANT: van Pee, Karl-Heinz  
 APPLICANT: Kliner, Sabine  
 TITLE OF INVENTION: Genes for the synthesis of  
 NUMBER OF INVENTION: antipathogenic substances  
 NUMBER OF SEQUENCES: 27  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ciba-Geigy Corporation  
 STREET: 520 White Plains Road  
 CITY: Tarrytown  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10591  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/729,214  
 FILING DATE: TBA  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Melis, J. Timothy  
 REGISTRATION NUMBER: 38,241  
 REFERENCE/DOCKET NUMBER: CGC 1506/CIP5  
 TELEPHONE: 919-541-8587  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 28958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-729-214-6

Query Match 3.5%; Score 43.6; DB 2; Length 28958;  
 Best Local Similarity 44.8%; Pred. No. 0.097;  
 Matches 290; Conservative 0; Mismatches 354; Indels 4; Gaps 3;

OY 148 GGCATGAGTGTCTCATCCAGCAGTGGCGGAGAGGGCTGCGTATACCGAGCGGGAT 207  
 DB 25306 GCGCCAGAGACCTTACCTCCGCGATGCGCTATTGCTGTGGCAACCGCGCCGAC 25365  
 OY 208 TTCAAGGGGCGAGCGCGCAACTGTGTGGCACCGGACCAAGTGTGGCGCGACCTGAT 267  
 DB 25366 TTGAGAACCGGGCGGCTCTGTAGCCCAACCGGAGAGAGCTCTCTCGCGGTGAC 25425  
 OY 268 TTATTGCTCAAGTCAA-AGAACCATAGGGCGGAAATACGGCCGCTTCCGACAGGGCG 326  
 DB 25426 TCGCTCGCCGACGAGAACCGCCCGGACACCGTCTCGAGGAGGAGGAGCCACGAC 25485  
 OY 327 ATCTGTTCACGTTCTTGCATTTGGCCGCGGCTGACGTGCTTGCACCATGCTTGTGAT 386  
 DB 25486 AAG-TGCTGTGCTTCTTCTGCGGCGAAGGCTGCGAGTGGGAAGGATGGCCCTTCTG 25545  
 OY 387 TCCGACACCACTCAATTTGCTTACGAGACCGTCCAGACCGCGGAGCGGACACTACCCCTG 446  
 DB 25546 CTCGACTCTCTGCGCGCTTCTTCTGCGGTCTGAGTCAAGATGCGAGCGCGGCTGTCT 25605  
 OY 447 CTTGCGCCCGATGAGCGAAGTGTGCGGTGACTGCGCGCCGCGGAGGTTGGCGCTTACAC--C 504  
 DB 25606 CACGTCGAGTGAAGCTGTGCTCTCTGCGCGCGGACAGTCAAGGGGCGCGCCCTCTCTG 25665  
 OY 505 TGAATCGAAGCCCAAGGGGCGCGGTGTGCTGATGAGGCGGGGTCTCCGCGGTGAACCG 564  
 DB 25666 CCGCTCGAGTGTACAGCGCGCGCTCTTGGCGCTCATGTGCTCTCTGGGCGGCGCTCTG 25725







FILING DATE: 23-DEC-1993  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 18740/135/CACO  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202 672 5300  
 TELEFAX: 202 672 5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2185 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 531..2069  
 FEATURE:  
 NAME/KEY: sig\_peptide  
 LOCATION: 531..902  
 FEATURE:  
 NAME/KEY: mat\_peptide  
 LOCATION: 903..2069  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 531..533  
 OTHER INFORMATION: /note="Met at position -124"  
 OTHER INFORMATION: represents fmet"  
 US-08-173-508-3

Query Match 3.4%; Score 42.4; DB 1; Length 2185;  
 Best Local Similarity 47.4%; Pred. No. 0.093;  
 Matches 127; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 524 CCGCGGTGCTGATGGCGGGGTCGCCGCGTCAACCGGCCGACGTCGTGATCGG 583  
 DB 1250 CGGCGTGGCATGACCGCGGTGGCCACCGGGGTGAAGGTGCAAGGTCTCCAA 1309  
 QY 584 CGCGGACACCGCGCGGTCAACGACGCGCGCATCGCCAAAGCATGGCGCGACGTTAC 643  
 DB 1310 CCGCGACGCGCTTCTTCTACACGAGCGCGGTGTCGCGGCTTCAATGTCGGCGGACCA 1369  
 QY 644 GGTCTAGACATCAACATGCAAACTTCGCAACTGACCGCGGAGTTCTGCGCGGAT 703  
 DB 1370 CGGCGTGGCATGACCAACAGCTATTACACGACCGCGTGTACTTCACTGCAAGGA 1429  
 QY 704 CCACACTGCTACTGACGCGGCTGAGAGTCCGTCAAAGTGGCGACCTGCT 763  
 DB 1430 CGACCGGACCAAGAGGCTGTCGAGGCGGCTTCGCGGCGCTCCGCTACGCGGAGAA 1489  
 QY 764 GATTGGGCGCTCTGTCGTCGACGAGCGCC 791  
 DB 1490 GAAGGCGCGGTCAACGTCGCGCGCGCC 1517

RESULT 15  
 US-08-265-310-3  
 Sequence 3, Application US/08265310  
 Patent No. 5856166  
 GENERAL INFORMATION:  
 APPLICANT: Bartfield, Daniel  
 APPLICANT: Butler, Michael J.  
 APPLICANT: Hadary, Dany  
 APPLICANT: Jemish, David  
 APPLICANT: Krieger, Timothy  
 APPLICANT: Malek, Lawrence T.  
 APPLICANT: Soostmeyer, Gisela  
 APPLICANT: Malczyk, Rya  
 APPLICANT: Kraysman, Phyllis  
 APPLICANT: Garven, Shelia

TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED  
 TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND  
 TITLE OF INVENTION: POLYPEPTIDES  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W.  
 CITY: Washington, D.C.  
 COUNTRY: USA  
 ZIP: 20007-5109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/265,310  
 FILING DATE: 24-JUN-1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/173,508  
 FILING DATE: 23-DEC-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 18740/133/CACO  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202 672 5300  
 TELEFAX: 202 672 5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2185 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 531..2069  
 FEATURE:  
 NAME/KEY: sig\_peptide  
 LOCATION: 531..902  
 FEATURE:  
 NAME/KEY: mat\_peptide  
 LOCATION: 903..2069  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 531..533  
 OTHER INFORMATION: /note="Met at position -124"  
 OTHER INFORMATION: represents fmet"  
 US-08-265-310-3

Query Match 3.4%; Score 42.4; DB 3; Length 2185;  
 Best Local Similarity 47.4%; Pred. No. 0.093;  
 Matches 127; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 524 CCGCGGTGCTGATGGCGGGGTCGCCGCGTCAACCGGCCGACGTCGTGATCGG 583  
 DB 1250 CGGCGTGGCATGACCGCGGTGGCCACCGGGGTGAAGGTGCAAGGTCTCCAA 1309  
 QY 584 CGCGGACACCGCGCGGTCAACGACGCGCGCATCGCCAAAGCATGGCGCGACGTTAC 643  
 DB 1310 CCGCGACGCGCTTCTTCTACACGAGCGCGGTGTCGCGGCTTCAATGTCGGCGGACCA 1369  
 QY 644 GGTCTAGACATCAACATGCAAACTTCGCAACTGACCGCGGAGTTCTGCGCGGAT 703  
 DB 1370 CGGCGTGGCATGACCAACAGCTATTACACGACCGCGTGTACTTCACTGCAAGGA 1429  
 QY 704 CCACACTGCTACTGACGCGGCTGAGAGTCCGTCAAAGTGGCGACCTGCT 763  
 DB 1430 CGACCGGACCAAGAGGCTGTCGAGGCGGCTTCGCGGCGCTCCGCTACGCGGAGAA 1489

Fri Jun 23 09:31:39 2000

us-09-362-485-8.mli

Page 12

Qy 764 GATTGGGGCCGCTGCTGTGCCAGGCGCC 791  
 Db 1490 GAAGGGCGCGGTCAACGTGCGCGGCGCC 1517

Search completed: June 22, 2000, 15:16:28  
Job time: 18466 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:07:49 ; Search time 5541.94 Seconds  
(without alignments)  
903.245 Million cell updates/sec

Title: US-09-362-485-8  
Perfect score: 1235  
Sequence: 1 ATCTTCAGATGATGCGAAC.....GGGAGCGATGATGCGGCC 1235

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues  
Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES

No.	Score	Match	Length	DB	ID	Description
C 1	57.4	4.6	925	82	CNS0091P	AL05013 Drosophila
C 2	54	4.4	925	82	CNS0091P	AL05013 Drosophila
C 3	51.8	4.2	535	74	AV39050S	AV39050S
C 4	49	4.0	910	82	CNS0060N	AL065629 Drosophila
C 5	46.6	3.8	844	82	CNS0052P	AL056522 Drosophila
C 6	46.6	3.8	846	82	CNS0108J	AL099337 Drosophila
C 7	45.4	3.7	591	69	AM128786	AL099337 Drosophila
C 8	45.4	3.7	935	82	CNS006XK	AL066051 Drosophila
C 9	45.2	3.7	1101	83	CNS0175T	AL108460 Drosophila
C 10	45	3.6	315	51	AV1712257	AV1712257 Drosophila
C 11	45	3.6	846	82	CNS0108J	AL099337 Drosophila
C 12	45	3.6	1101	83	CNS0181E	AL099337 Drosophila
C 13	44.8	3.6	932	82	CNS0072O	AL066051 Drosophila
C 14	44.2	3.6	932	82	CNS006XK	AL066051 Drosophila
C 15	43.8	3.5	427	51	AV1736678	AV1736678 Drosophila
C 16	43.8	3.5	512	74	AV36794	AV36794 Drosophila
C 17	43.8	3.5	521	74	AV397116	AV397116 Drosophila
C 18	43.8	3.5	526	74	AV394827	AV394827 Drosophila
C 19	43.8	3.5	744	83	CNS0172K	AL107126 Drosophila
C 20	43.6	3.5	1101	83	CNS016UE	AL066051 Drosophila
C 21	43.4	3.5	414	80	AM286084	AM286084 Drosophila
C 22	43.4	3.5	465	80	AM287007	AM287007 Drosophila
C 23	43.2	3.5	439	91	AO864063	AO864063 Drosophila
C 24	43.2	3.5	798	82	CNS000AJA	AL055851 Drosophila
C 25	43	3.5	413	74	AV396947	AV396947 Drosophila
C 26	42.6	3.4	645	82	CNS01213	AL101589 Drosophila
C 27	42.4	3.4	843	82	CNS00C51	AL059666 Drosophila
C 28	41.8	3.4	530	74	AV396353	AV396353 Drosophila
C 29	41.8	3.4	932	82	CNS0072O	AL066051 Drosophila
C 30	41.6	3.4	506	74	AM180713	AM180713 Drosophila
C 31	41.6	3.4	870	82	CNS006E2	AL064271 Drosophila
C 32	41.4	3.4	1101	82	CNS00EXJ	AL078875 Drosophila
C 33	41.2	3.3	639	60	AI812194	AI812194 Drosophila
C 34	41.2	3.3	774	105	AO327321	AO327321 Drosophila
C 35	41.2	3.3	910	82	CNS0060N	AL056529 Drosophila
C 36	41	3.3	605	46	AI461529	AI461529 Drosophila
C 37	41	3.3	1101	83	CNS0175T	AL108460 Drosophila
C 38	40.8	3.3	384	64	AM054773	AM054773 Drosophila
C 39	40.8	3.3	385	21	D48746	D48746 RCSI5137A
C 40	40.8	3.3	426	40	C91740	C91740 Rice
C 41	40.8	3.3	437	74	AV395790	AV395790 Drosophila
C 42	40.8	3.3	839	82	CNS004NB	AL055280 Drosophila
C 43	40.8	3.3	964	82	CNS003MG	AL065254 Drosophila
C 44	40.6	3.3	506	74	AV396405	AV396405 Drosophila
C 45	40.6	3.3	574	91	AO848186	AO848186 LMAFW1_1

## ALIGNMENTS

RESULT	1				
CNS0091P/C					
LOCUS	CNS0091P	925 bp	DNA	GSS	03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR16 of Rpcl-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL053013				
VERSION	AL053013.1	GI:4934461			
KEYWORDS	GSS.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscivora; Ephydroidea; Drosophilidae; Drosophila.				
AUTHORS	1 (bases 1 to 925)				
TITLE	Genoscope.				
JOURNAL	Direct Submission				
COMMENT	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
	Determination of this BAC-end sequence was carried out as part of a				

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org>. The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Oseegawa and Aaron Mammoler in Peter de Jong's Laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2: cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPC Resource Center can be found at <http://bacpac.med.buffalo.edu/drosophila.bac.htm>.

```
FEATURES
source      Location/Qualifiers
1..925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="PBC1-98"
/clone="BACR19D16"
/note="end : 1973"

BASE COUNT      120 a      61 c      61 g      172 t      511 others

ORIGIN
```

Query Match	4.68;	Score 57.4;	DB 82;	Length 925;
Best Local Similarity	11.9%;	Pred. NO. 0.057;		
Matches	43;	Conservative 176;	Mismatches 142;	Indels 0;
			Gaps	0;

QY	441	CCCCCTGCTGCCCGCATGAGCGAAGTGCCTGGCTGACTCCCGCCACAGTTGGCCTTAC	500
Db	920	SCSCSCSBSBSSSSSMSTSSNSBSCSSBSBSTSSMSSSBSSSSSGSSSSSTGS	861
QY	501	CACCTGATGCACCCCAGAAGGCGCCCGGGTGTCGTATGAGCGGGGTGCCGGCGTCGA	560
Db	860	SACVKNAMSSCGCCCGCGMABCMCSSSSSCCSASRKYVRASGGAKRGGSAGASH	801
QY	561	CCGAGCCGACGTGTCGTATCGCGCCGCGCACCCCGCTACAACAGCAGCCCGCATCGCC	620
Db	800	SHSSSSACBSSSSSCSASCMSASSSSSSASRSRSGGAGSGAGSSSRSSSSSASAG	741
QY	621	AACGGCATGGCGCGGACCCTTAGCGTTCTAGACTCAACATCGACAACCTGGCGAACTC	680
Db	740	VVSSASSSSSSSCSSSVACSSVASMSCSSBSBSASASSSSSSASASCASCOCCTTSWC	681
QY	681	GACCCGAGTCTCGCGCGCGATCCACACTCGCTACTCATCGGCTACGAGTCGAGGGT	740
Db	680	SCRTSAMSAARSSSSSSSSSCSSSMASASSASSSSSSSSSSSSSGSACGBMSG	621
QY	741	GCGCTCAAACCTGCCGATCTTCGATTGGGCGCTCTCTGTGTCGAGCGCCGAAGCACCC	800
Db	620	GGSGSVASMSGMSSSVSNSGGRSGSGCGGVGDSGSSSGSGSGSGSVSCSSGCMK	561
QY	801	A	801
Db	560	R	560

RESULT	2
CNS00091p	
LOCUS	
DEFINITION	CNS00091p 925 bp DNA GSS 03-JUN-1999
ACCESSION	Drosophila melanogaster genome survey sequence TET3 end of BAC #
VERSION	BACR19b16 of RPc1-98 library from Drosophila melanogaster (fruit
KEYWORDS	fly). genomic survey sequence.
SOURCE	AL053013
ORGANISM	AL053013.1 GI:4934461
REFERENCE	GSS.
	fruit fly.
	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
	1 (bases 1 to 925)



Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

## source

Location/Qualifiers  
1. .910  
/organism="Drosophila melanogaster"

/db.xref="taxon:7227"  
/clone.lib="RPCI-98"  
/clone="BACR14J21"

/note="end : 177"

BASE COUNT 202 a 63 c 112 g 198 t 335 others  
ORIGIN

Query Match 4.0% Score 49; DB 62; Length 910;  
Best Local Similarity 20.5% Pred. No. 2.4; Mismatches 85; Indels 0; Gaps 0;  
Matches 47; Conservative 97; Mismatches 85; Indels 0; Gaps 0;

QY 392 CACCACTGATGCTGCTAGAGACCGCTCCAGACGCCGGAAGCGCACTACCCCTGCTGC 451  
Db 713 SRVMSAGSCTASSSSSSASASSASASASASASASASASASASASASASASASASAS 654  
QY 452 CCCGATGAGCGAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 511  
Db 653 VMAVVAASSSSVASGSSSVSCSSRCCASSSASASASASASASASASASASASASASAS 594  
QY 512 AACCCAGAGGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 571  
Db 593 CASSSSRSGSCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 534  
QY 572 CCGTGTGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 620  
Db 533 CASCSCGCGCCSSSCSSCCSCCCSCAMAMAHACCCSCCCSCCCDC 485

## RESULT 5

CNS0052P 844 bp DNA GSS 03-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #  
DEFINITION BACR14J21 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
AL056652.1 GI:4932342

ACCESSION AL056652.1 GI:4932342  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 844)  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osogawa and  
Aaron Mammot in Peter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

Location/Qualifiers  
1. .844  
/organism="Drosophila melanogaster"

/db.xref="taxon:7227"  
/clone.lib="RPCI-98"  
/clone="BACR14J21"

BASE COUNT 261 a 112 c 92 g 35 t 344 others  
ORIGIN

Query Match 3.8% Score 46.6; DB 82; Length 844;  
Best Local Similarity 15.8% Pred. No. 6.8; Mismatches 146; Indels 0; Gaps 0;  
Matches 58; Conservative 146; Mismatches 146; Indels 0; Gaps 0;

QY 886 AAGCTCAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 945  
Db 733 MMTTCVMMVMMVMMVMMVMMVMMVMMVMMVMMVMMVMMVMMVMMVMMVMMVMMVMM 674  
QY 946 ACTGCTGCGGACATGCG 1005  
Db 673 TTTNNTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTTMTT 614  
QY 1006 CGAGATCCGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1065  
Db 613 SSS 554  
QY 1066 ATCCGCGACTACCAAGGCTTTTTCGACGACGACGACGACGACGACGACGACGACG 1125  
Db 553 SSSSBBCSSBSSCCG 494  
QY 1126 CCACGACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1185  
Db 493 GKSCKGKGGKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGKSG 434  
QY 1186 CTCGTTACCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1235  
Db 433 GSSSKTSSSSSCBSSSSSTGSSGSKKYSYSTSSSSSCGCTGKNTTTCV 384

## RESULT 6

CNS010RJ 846 bp DNA GSS 26-JUL-1999  
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC  
DEFINITION BACN04N13 of DrosPAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
AL099337.1 GI:5619448

ACCESSION AL099337.1 GI:5619448  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 846)  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC  
library (DrosBAC) was made by Alain Billand at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
Project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBeloBAC11.  
Location/Qualifiers  
1. .846  
/organism="Drosophila melanogaster"









QY	388	CCGGACACACGTCATTCTCTACGAGACCGCCGGAAGGGCCACTACCTCG	44.7
	11	: : : : :     : : : : :     : : : : :     : : : : :	
Db	308	CCCCCCTVYMYVMGMAAAACCCCCCAVNASCSGCCSCSCSSCSGCSGCSG	24.9
QY	448	TTGCCCCGATGACGCAAGTCGCCCGGTGCATCGCCGCCCAAGTTGGCGCTTACACCTGA	50.7
	11	: : : : :     : : : : :     : : : : :     : : : : :     : : : : :	
Db	248	AAGCGMSMGVGGCAACAAGGGCCGAGAACCCSCCACCSCCCSSSCGSCSSCSCSSSS	18.9
QY	508	TGCGAATCCCAAGGGGGCCCGGGTGTGCTGATGGGGCGGGGTGCCCGGGGTGGAACCCGGCG	56.7
	11	:     : : : : :     : : : : :     : : : : :	
Db	188	CGCGACCCSSSAGGGGCCSSSBCGCCCCACCCCGCCSGACVGAACCGCCAC -CMYVGGGSG	13.0
QY	568	ACGTGCTGTGATCGGCGCCGCGCACCCGCGGTCTCAACGCAAGCCCGCATGCGCAACGGCA	62.7
	11	: : : : :     : : : : :     : : : : :     : : : : :	
Db	129	ACGGAAAGGAGAGAGGGCGCGGNACCCCGCCACACCCCGCCCGCCGACAGCGCGCGGCA	7.0
QY	628	TGGGGCGCGACG	63.8
	11		
Db	69	AGCGCCCAACG	5.9

LOCUS	DEFINITION	FEATURES
CNS0072Q	932 bp	DNA
CNS0072Q	Drosophila melanogaster genome survey sequence T7 end of BAC # BRORIAB09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.	GSS

ACCESSION	AL066742
VERSION	AL066742.1
	GI:4945205

KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	fruit fly.

**ORGANISM**

## REFERENCE

AUTHORS  
TITLE  
JOURNAL

**COMMENT**

Submitted (22-JUN-1999) Genoscope - Centre National de Séquençage  
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
Web : www.genoscope.cns.fr)  
Determination of this *Drosophila* Genome Project (BDGP).  
Collaboration with the Berkeley *Drosophila* Genome Project (BDGP).  
The BDGP is constructing a physical map of the *Drosophila*  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP *Drosophila*  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mammosser in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPc1-98 and was constructed by partial  
EcoRI digestion of *Drosophila* DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp. the same strain used for the BDGP's  
p1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

**Source**

BASE COUNT	155 a	202 c	241 g	91 t	243 others
ORIGIN					

Query Match 3.6%; Score 44.8; DB 82; Length 932;  
 Best Local Similarity 34.1%; Pred.No.15;  
 Matches 73; Conservative 47; Mismatches 94; Indels 0;  
 Gaps 0.  
 432 GCGGACGTAACCCCTCCTTCCCCGATGAGCCAGTCGCGGTCGACTGCCCCCAAGTT 491

[illegible]

DB 859

RESULT 14  
CNS005YK/5

LOCUS	935 bp	DNA	GSS	03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC			

BCR14N09 of RPI-98 library from *Drosophila melanogaster* (fruit fly), genomic survey sequence.

VERSION AL066051.1 GI:4945019

KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
1 (bases 1 to 935)

**AUTHORS**  
**TITLE**  
**JOURNAL**

COMMENT  
BP - 01 91006 EVRY cedex - FRANCE (E-mail : [segrete@genoscope.cns.fr](mailto:segrete@genoscope.cns.fr)  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))  
Determination of this BAC-end sequence was carried out as part of a

determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see: [http://www.fruitfly.org/TheBDGP/Drosophila\\_melanogaster\\_BAC\\_library](http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library) was prepared by Kazutoyo Oseegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

**FEATURES**  
**SOURCE**

Source

BASE COUNT	257 a	170 c	162 g	96 t	250 others
ORIGIN					

Query Match	3.68;	Score 44.2;	DB 82;	Length 935;
Best Local Similarity	27.48;	Pred. No. 20;		
Matches 93;	Conservative 83;	Mismatches 163;	Indels 0;	Gaps 0

406 CCTACGAGACCGTCCAGACCGCCGAAGGGCCACTAACCCCTGCTTGCCCCGATGAGCGAAG 465

[illegible]

466 TCGCCGCTGACTCGCCGCCAGGTGGCGCTTACCACCTGATGCGAACCACAAGGGGCC 525

Db 796 5GGCCSGGCGSSGGCGCGCCGSSCCGCGSKSKCGCGCGSGGSSSSSSGSGCGSGGS 737





```

RESULT 1
LOCUS A87612 1209 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 9 from Patent W09836089.
ACCESSION A87612
VERSION A87612.1 GI:6736252
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1209)
AUTHORS Flohe, L. and Singh, M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
source
Location/Qualifiers
1..1209
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 228 a 391 c 373 g 216 t 1 others
ORIGIN

```

```

Query Match 99.9%; Score 1208; DB 5; Length 1209;
Best Local Similarity 100.0%; Pred. No. 1,1e-165;
Matches 1209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ATCTGCAGATTAATCAACTTTCTTACACTGAAGGTACAGTATCGAAGGGGTATC 60
DB 1 ATCTGCAGATTAATCAACTTTCTTACACTGAAGGTACAGTATCGAAGGGGTATC 60
QY 61 ATGGCGCTCGGTTATTCGACGAGACCAAAAACAAAGAAATTCGGGTGCGCATCACCCG 120
DB 61 ATGGCGCTCGGTTATTCGACGAGACCAAAAACAAAGAAATTCGGGTGCGCATCACCCG 120
QY 121 GCCGCGCTCGGGAAGTACACCGCTGTGGCCATGAGGTGCTCATCCAGGCAAGTGGCGGA 180
DB 121 GCCGCGCTCGGGAAGTACACCGCTGTGGCCATGAGGTGCTCATCCAGGCAAGTGGCGGA 180
QY 122 GCCGCGCTCGGGAAGTACACCGCTGTGGCCATGAGGTGCTCATCCAGGCAAGTGGCGGA 180
DB 122 GCCGCGCTCGGGAAGTACACCGCTGTGGCCATGAGGTGCTCATCCAGGCAAGTGGCGGA 180
QY 181 GAGGCGCTCGGTTATTCAGGACGCGGATTTCAAGGCGGCAAGCTGCTGGGCAACC 240
DB 181 GAGGCGCTCGGTTATTCAGGACGCGGATTTCAAGGCGGCAAGCTGCTGGGCAACC 240
QY 241 GCCGACGAGGTGTGGGCGGACGCTGATTTATGCTCAAGGCAAGAAACGATAGCGGC 300
DB 241 GCCGACGAGGTGTGGGCGGACGCTGATTTATGCTCAAGGCAAGAAACGATAGCGGC 300
QY 301 GAATACGCGCGCTGCGACACGCGGCAATCTGTTCAAGTTCCTTCAATTGGCCGCTCA 360
DB 301 GAATACGCGCGCTGCGACACGCGGCAATCTGTTCAAGTTCCTTCAATTGGCCGCTCA 360
QY 361 CGTCTTGCACCGATGGTGTGTTGATTCCGGGCAACGTAATTTGGCTACGAGACGCTC 420
DB 361 CGTCTTGCACCGATGGTGTGTTGATTCCGGGCAACGTAATTTGGCTACGAGACGCTC 420
QY 421 CAGACCGCGGACGCGGCACTACCCCTGCTGCCCCGATGAGCGGAAGTCCGCGTCACTC 480
DB 421 CAGACCGCGGACGCGGCACTACCCCTGCTGCCCCGATGAGCGGAAGTCCGCGTCACTC 480
QY 481 GCCGCGGAGGTGTGGCGCTTACCACTGATGGAACCCAAAGGGGCGCGGTGCTGATG 540
DB 481 GCCGCGGAGGTGTGGCGCTTACCACTGATGGAACCCAAAGGGGCGCGGTGCTGATG 540
QY 541 GCGCGGGGTGCGCGGCTGGAACCGCGGAGCTGTGGTATCGGCGCGGACCGCGGC 600
DB 541 GCGCGGGGTGCGCGGCTGGAACCGCGGAGCTGTGGTATCGGCGCGGACCGCGGC 600
QY 601 TACAACGACGCGGCTATCGCAACGGGATGGGCGGACCGTTCAGGTTAGACATCAAC 660
DB 601 TACAACGACGCGGCTATCGCAACGGGATGGGCGGACCGTTCAGGTTAGACATCAAC 660
QY 661 ATGCACAAACTTGGGCACTGACGCGGAGTTCGTGGCGCGGATTCACACTCGCTACTCA 720

```

```

DB 661 ATGCACAAACTTGGGCACTGACGCGGAGTTCGTGGCGCGGATTCACACTCGCTACTCA 720
QY 721 TCGGCTACGAGCTCGAGGCTGCCGTCAAACGTCGCGACCTGTGTATGGGCGCTCTG 780
DB 721 TCGGCTACGAGCTCGAGGCTGCCGTCAAACGTCGCGACCTGTGTATGGGCGCTCTG 780
QY 781 GTGGCGGCGGCGCAAGGCAACCAATAGTCTGGAATTCATCTGTGGCGATATGAACCA 840
DB 781 GTGGCGGCGGCGCAAGGCAACCAATAGTCTGGAATTCATCTGTGGCGATATGAACCA 840
QY 841 GTGGCGGCTACTGTGATATATGATATGACATGACACGAGGCGGCTGTTGAAAGCTCACAGCG 900
DB 841 GTGGCGGCTACTGTGATATATGATATGACATGACACGAGGCGGCTGTTGAAAGCTCACAGCG 900
QY 901 ACCACCTAGACACCCGACGCTTCCGCTGACAGACACGCTGTTTACTGTGGCGGAAC 960
DB 901 ACCACCTAGACACCCGACGCTTCCGCTGACAGACACGCTGTTTACTGTGGCGGAAC 960
QY 961 ATGCGCGGCTCGGTTGCGGAAGAGTGAAGTACGCGCTGACCAAGCGGAGTGGCGTAT 1020
DB 961 ATGCGCGGCTCGGTTGCGGAAGAGTGAAGTACGCGCTGACCAAGCGGAGTGGCGTAT 1020
QY 1021 GTGCTGAGAGCTTCCGACCATGAGCTGGCGGCGGCTGCGGCTGCAATCCGACATAAGCC 1080
DB 1021 GTGCTGAGAGCTTCCGACCATGAGCTGGCGGCGGCTGCGGCTGCAATCCGACATAAGCC 1080
QY 1081 AAAGGTCTTTGCAAGCGACGCAAGGGGCGTTACTGTCCGAACGGGTGGCCACCACTGGGG 1140
DB 1081 AAAGGTCTTTGCAAGCGACGCAAGGGGCGTTACTGTCCGAACGGGTGGCCACCACTGGGG 1140
QY 1141 GTGCGCTTACCGACGCGGCGGACGCTGTGCGCTGACTCTCGGCGGCTGTTAGCGCGAG 1200
DB 1141 GTGCGCTTACCGACGCGGCGGACGCTGTGCGCTGACTCTCGGCGGCTGTTAGCGCGAG 1200
QY 1201 CACGCTCG 1209
DB 1201 CACGCTCG 1209

```

```

RESULT 2
LOCUS A87613 1236 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 10 from Patent W09836089.
ACCESSION A87613
VERSION A87613.1 GI:6736253
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1236)
AUTHORS Flohe, L. and Singh, M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
source
Location/Qualifiers
1..1236
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 395 c 385 g 220 t
ORIGIN

```

```

Query Match 99.9%; Score 1208; DB 5; Length 1236;
Best Local Similarity 99.9%; Pred. No. 1,1e-165;
Matches 1208; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 ATCTGCAGATTAATCAACTTTCTTACACTGAAGGTACAGTATCGAAGGGGTATC 60
DB 1 ATCTGCAGATTAATCAACTTTCTTACACTGAAGGTACAGTATCGAAGGGGTATC 60
QY 61 ATGGCGCTCGGTTATTCGACGAGACCAAAAACAAAGAAATTCGGGTGCGCATCACCCG 120

```

```

Db      61 ATGGCGGTGGTATTTCGACCGGACCAAAAACGAATTCGGGGTGGGCATCACCCG 120
QY      121 GCCGGCGTGGGAACTTAACCCGTCGTGGCATGAGTGTCTATCCAGCAGTGGCCGA 180
Db      121 GCCGGCGTGGGAACTTAACCCGTCGTGGCATGAGTGTCTATCCAGCAGTGGCCGA 180
QY      181 GAGGGGTGGGCTATCCGACCGGAGATTCGAAAGGCGGAGGCGGACGACGTGGCGCACC 240
Db      181 GAGGGGTGGGCTATCCGACCGGAGATTCGAAAGGCGGAGGCGGACGACGTGGCGCACC 240
QY      241 GCGGACCAAGTGTGGGCGGACGCTGATTTATTCCTCAAGGTCAAAAGACCGATAGCGGG 300
Db      241 GCGGACCAAGTGTGGGCGGACGCTGATTTATTCCTCAAGGTCAAAAGACCGATAGCGGG 300
QY      301 GAATACGCGCCGCTGGGACACGGGCGAGATCTTTCACGTTCTTGCAATTTGGCCGCTA 360
Db      301 GAATACGCGCCGCTGGGACACGGGCGAGATCTTTCACGTTCTTGCAATTTGGCCGCTA 360
QY      361 CGTGGTTGACCGGATCGTTGTGATTCGGGACGACGCAATTCGCTACGAGACCGTC 420
Db      361 CGTGGTTGACCGGATCGTTGTGATTCGGGACGACGCAATTCGCTACGAGACCGTC 420
QY      421 CAGACCGCGGACGCGGCGGACGCTGCTGCTGCGGATG  GCGAAGTCGCGGCTCACTC 480
Db      421 CAGACCGCGGACGCGGCGGACGCTGCTGCTGCGGATG  GCGAAGTCGCGGCTCACTC 480
QY      481 GCGGCGGAGTTGGGCGCTTACCACTGATGCGGAACCCAAAGGGGGCGCGGTGGCTGATG 540
Db      481 GCGGCGGAGTTGGGCGCTTACCACTGATGCGGAACCCAAAGGGGGCGCGGTGGCTGATG 540
QY      541 GCGGCGGAGTTGGGCGCTTACCACTGATGCGGCGGAGAGTGTGATGCGGCGGCGGCGG 600
Db      541 GCGGCGGAGTTGGGCGCTTACCACTGATGCGGCGGAGAGTGTGATGCGGCGGCGGCGG 600
QY      601 TACACGACGACCGCGCATCGGCAAGCGCATGCGGCGGACGCTTACGATGATGATGATG 660
Db      601 TACACGACGACCGCGCATCGGCAAGCGCATGCGGCGGACGCTTACGATGATGATGATG 660
QY      661 ATGGAACAACTTGGGAACTCGAGCGCGGAGTTGTGGGCGGATCCACATCTGCTACTCA 720
Db      661 ATGGAACAACTTGGGAACTCGAGCGCGGAGTTGTGGGCGGATCCACATCTGCTACTCA 720
QY      721 TCGGCTTACGAGTTCGAGGGTGGCGTCAAAAGTGGCGGACGCTGATGATGAGGCGCTCTG 780
Db      721 TCGGCTTACGAGTTCGAGGGTGGCGTCAAAAGTGGCGGACGCTGATGATGAGGCGCTCTG 780
QY      781 GTGCGAGGCGGCGGACGCAACCAATTAATGTCGAATTCACCTTGTGCGGATATGAACCA 840
Db      781 GTGCGAGGCGGCGGACGCAACCAATTAATGTCGAATTCACCTTGTGCGGATATGAACCA 840
QY      841 GGTGGGCTACTGTTGATATAGCCATCGACAGGAGGGGCGTTCGAAGGCTACAGCGG 900
Db      841 GGTGGGCTACTGTTGATATAGCCATCGACAGGAGGGGCGTTCGAAGGCTACAGCGG 900
QY      901 ACCACTTACGACACCGGAGTTCGCGGTCGACGACGAGTGTGTTTACTGCTGGGGAAC 960
Db      901 ACCACTTACGACACCGGAGTTCGCGGTCGACGACGAGTGTGTTTACTGCTGGGGAAC 960
QY      961 ATGCGCGGCTCGGTGCGGAGAGAGTGCACCTTACGCGCTGACCAACGCGAGATGCCAT 1020
Db      961 ATGCGCGGCTCGGTGCGGAGAGAGTGCACCTTACGCGCTGACCAACGCGAGATGCCAT 1020
QY      1021 GTGCTGAGGTTTCGACCAATGCTGCTGGGCGGCGGCTGCGGCTGCAATTCGGGCACTAGCC 1080
Db      1021 GTGCTGAGGTTTCGACCAATGCTGCTGGGCGGCGGCTGCGGCTGCAATTCGGGCACTAGCC 1080
QY      1081 AAGGCTTTTCGACGACGAGAGGGGGCTTACTTTCGAAAGGGGTGGCGACGACTGGGG 1140
Db      1081 AAGGCTTTTCGACGACGAGAGGGGGCTTACTTTCGAAAGGGGTGGCGACGACTGGGG 1140
QY      1141 GTGCGGTTACGAGCGCGGCGGAGGCTGTGCTGACTCTCGGCGGCTGTTACGCCGAG 1200
Db      1141 GTGCGGTTACGAGCGCGGCGGAGGCTGTGCTGACTCTCGGCGGCTGTTACGCCGAG 1200

```

```

QY      1201 CACAGCTCG 1209
Db      1201 CACAGCTCG 1209

RESULT 3
A89753
LOCUS      A89753          1236 bp      DNA
DEFINITION Sequence 10 from Patent WO90,2862.
ACCESSION A89753
VERSION    A89753.1 GI:6738287
KEYWORDS
SOURCE
ORGANISM   unidentified.
            unidentified.
AUTHORS    Flohe L. and Singh M.
TITLE      L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL    Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
            source
            location/Qualifiers
            1..1236
            /organism="unidentified"
            /db_xref="taxon:32644"

BASE COUNT 236 a 395 c 385 g 220 t
ORIGIN
Query Match          99.9% Score 1208; DB 5; Length 1236;
Best Local Similarity 99.9%; Pred. No. 1..1e-165;
Matches 1208; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ACTTTCAGATTAAATGAACTTCTTACACTGAAAGCGTACGATTCGAGAGGGTAATC 60
Db      1 ACTTTCAGATTAAATGAACTTCTTACACTGAAAGCGTACGATTCGAGAGGGTAATC 60
QY      61 ATGCGGTGGGTATTCGCGACCGAGACCAAAACAAAGAAATTCGGGTGGCCATCACCCG 120
Db      61 ATGCGGTGGGTATTCGCGACCGAGACCAAAACAAAGAAATTCGGGTGGCCATCACCCG 120
QY      121 GCGGCGTGGCGGAACTA..CCGTCGTGGCATGAGGTGCTCATCCAGGACGATGCCGGA 180
Db      121 GCGGCGTGGCGGAACTA..CCGTCGTGGCATGAGGTGCTCATCCAGGACGATGCCGGA 180
QY      181 GAGGGCTCGGCTATACCGGACGCGGATTTCAAGGGCGGACGCGCACTGGTCGGCACC 240
Db      181 GAGGGCTCGGCTATACCGGACGCGGATTTCAAGGGCGGACGCGCACTGGTCGGCACC 240
QY      241 GCGGACGAGTGTGGGCGGCAAGCTGATTTATTCCTCAAGGTCAAAAGACGATAGCGCG 300
Db      241 GCGGACGAGTGTGGGCGGCAAGCTGATTTATTCCTCAAGGTCAAAAGACGATAGCGCG 300
QY      301 GAATACGCGCGGCTGGGACGCGGAGATCTGTCACGTTCTTGATTTGGCGCGCTCA 360
Db      301 GAATACGCGCGGCTGGGACGCGGAGATCTGTCACGTTCTTGATTTGGCGCGCTCA 360
QY      361 CGTGGTTGACCGGATCGTTGTGATTCGGGACGACGCTCAATTTGCTACAGAGACCGTC 420
Db      361 CGTGGTTGACCGGATCGTTGTGATTCGGGACGACGCTCAATTTGCTACAGAGACCGTC 420
QY      421 CAGACCGCGGAGGCGGACATACCCCTGCTTGGCCCGGATAGGGAAGTCCCGGTGAGCTC 480
Db      421 CAGACCGCGGAGGCGGACATACCCCTGCTTGGCCCGGATAGGGAAGTCCCGGTGAGCTC 480
QY      481 GCGGCGGAGTTGGGCGCTTACCACTGATGCGAACCAAGGGGGCGCGGATGCTGATG 540
Db      481 GCGGCGGAGTTGGGCGCTTACCACTGATGCGAACCAAGGGGGCGCGGATGCTGATG 540
QY      541 GCGGCGGAGTTGGGCGCTTACCACTGATGCGAACCAAGGGGGCGCGGATGCTGATG 540
Db      541 GCGGCGGAGTTGGGCGCTTACCACTGATGCGAACCAAGGGGGCGCGGATGCTGATG 540

```



QY 601 TACACGACGCGCGATGCGCAACGCGATGGCGGACCGCTTACGGTTACATCAAC 660  
 DB 601 TACACGACGCGCGATGCGCAACGCGATGGCGGACCGCTTACGGTTACATCAAC 660  
 QY 661 ATGACAACTTGGGCAATCGACGCGGATGCTGGCGGATCCACACTGCTACTCA 720  
 DB 661 ATGACAACTTGGGCAATCGACGCGGATGCTGGCGGATCCACACTGCTACTCA 720  
 QY 721 TCGGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCT 780  
 DB 721 TCGGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCT 780  
 QY 781 GTCGACGCGCGCAAGGCGCAATTAATGCTGCAATTAATGCTGCGGATGCAACCA 840  
 DB 781 GTCGACGCGCGCAAGGCGCAATTAATGCTGCAATTAATGCTGCGGATGCAACCA 840  
 QY 841 GTCGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCT 900  
 DB 841 GTCGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCT 900  
 QY 901 ACCACCTAGACGACCGGACGCTGGCGGATGCTGGCGGATGCTGGCGGATGCT 960  
 DB 901 ACCACCTAGACGACCGGACGCTGGCGGATGCTGGCGGATGCTGGCGGATGCT 960  
 QY 961 ATGCGCGCTGCTGGCGGATGCTGGCGGATGCTGGCGGATGCTGGCGGATGCT 1020  
 DB 961 ATGCGCGCTGCTGGCGGATGCTGGCGGATGCTGGCGGATGCTGGCGGATGCT 1020  
 QY 1021 GTCGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCT 1080  
 DB 1021 GTCGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCT 1080  
 QY 1081 AAGGCTTTCGACGACGCAAGGCGGATGCTGGCGGATGCTGGCGGATGCTGGCG 1140  
 DB 1081 AAGGCTTTCGACGACGCAAGGCGGATGCTGGCGGATGCTGGCGGATGCTGGCG 1140  
 QY 1141 GTCGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCT 1200  
 DB 1141 GTCGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCT 1200  
 QY 1201 CMCAGCTG 1209  
 DB 1201 CMCAGCTG 1209  
 RESULT 4  
 A87606 LOCUS A87606 1235 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 3 from Patent WO9836089.  
 ACCESSION A87606  
 VERSION A87606.1 GI:6736246  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1235)  
 FLOHE, L. and SINGH, M.  
 TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source 1..1235  
 Location/Qualifiers  
 BASE COUNT 236 a 394 c 382 g 220 t 3 other:  
 ORIGIN  
 Query Match 99.8%; Score 1207; DB 5; Length 1235;  
 Best Local Similarity 99.8%; Pred. No. 1 5e-155;  
 Matches 1207; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 ATCTTGACGATTAATCGAATTTCTTACACTGAGCGTACAGATGAGAGGGGTAAATC 60

DB 1 ATCTTGACGATTAATCGAATTTCTTACACTGAGCGTACAGATGAGAGGGGTAAATC 60  
 QY 61 ATGCGCGCTGGTATTCGACGAGACCAAAACAAAGATTCGCGGTGGCATACCCG 120  
 DB 61 ATGCGCGCTGGTATTCGACGAGACCAAAACAAAGATTCGCGGTGGCATACCCG 120  
 QY 121 GCCGCGCTGGGGAATTAACCGTGTGGCCTAGAGTGTCTATTCAGAGCAAGTGTCCGGA 180  
 DB 121 GCCGCGCTGGGGAATTAACCGTGTGGCCTAGAGTGTCTATTCAGAGCAAGTGTCCGGA 180  
 QY 181 GAGGCGCTGGGATACCGGAGCGGATTTCAAGGCGGACCGCAACTGGTGGCACC 240  
 DB 181 GAGGCGCTGGGATACCGGAGCGGATTTCAAGGCGGACCGCAACTGGTGGCACC 240  
 QY 241 GCCGACGAGCTGTGGGCGGACGCTGATTTATTTGCTCAAGGTCAAAGAACGATAGCGCG 300  
 DB 241 GCCGACGAGCTGTGGGCGGACGCTGATTTATTTGCTCAAGGTCAAAGAACGATAGCGCG 300  
 QY 301 GAATAGGCGCGCTGGGACGAGGCGGATTTGCTGACGCTTGGCATTTGGCGCGCA 360  
 DB 301 GAATAGGCGCGCTGGGACGAGGCGGATTTGCTGACGCTTGGCATTTGGCGCGCA 360  
 QY 361 GTCGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCTGGCG 420  
 DB 361 GTCGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCTGGCG 420  
 QY 421 CAGACCGCGGACGAGGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
 DB 421 CAGACCGCGGACGAGGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
 QY 481 GCCGCGCGGCTGGGCTTACACCTGATGCAACCAAGGCGGCGGCGGCTGCTGATG 540  
 DB 481 GCCGCGCGGCTGGGCTTACACCTGATGCAACCAAGGCGGCGGCGGCTGCTGATG 540  
 QY 541 GCGGCGGCTGGGCGGCTGCAACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
 DB 541 GCGGCGGCTGGGCGGCTGCAACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
 QY 601 TACACGACGCGCGGATGCGCAACGCGGATGCTGGCGGATGCTGGCGGATGCTGGCG 660  
 DB 601 TACACGACGCGCGGATGCGCAACGCGGATGCTGGCGGATGCTGGCGGATGCTGGCG 660  
 QY 661 ATGACAACTTGGGCAATCGACGCGGATGCTGGCGGATGCTGGCGGATGCTGGCG 720  
 DB 661 ATGACAACTTGGGCAATCGACGCGGATGCTGGCGGATGCTGGCGGATGCTGGCG 720  
 QY 721 TCGGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCTGGCG 780  
 DB 721 TCGGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCTGGCG 780  
 QY 781 GTCGACGCGCGCAAGGCGCAATTAATGCTGCAATTAATGCTGCGGATGCAACCA 840  
 DB 781 GTCGACGCGCGCAAGGCGCAATTAATGCTGCAATTAATGCTGCGGATGCAACCA 840  
 QY 841 GTCGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCTGGCG 900  
 DB 841 GTCGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCTGGCG 900  
 QY 901 ACCACCTAGACGACCGGACGCTGGCGGATGCTGGCGGATGCTGGCGGATGCTGGCG 960  
 DB 901 ACCACCTAGACGACCGGACGCTGGCGGATGCTGGCGGATGCTGGCGGATGCTGGCG 960  
 QY 961 ATGCGCGCTGCTGGCGGATGCTGGCGGATGCTGGCGGATGCTGGCGGATGCTGGCG 1020  
 DB 961 ATGCGCGCTGCTGGCGGATGCTGGCGGATGCTGGCGGATGCTGGCGGATGCTGGCG 1020  
 QY 1021 GTCGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCTGGCG 1080  
 DB 1021 GTCGCTTACGAGCTGAGGGTGGCTGCAACAGTGGCGGATGCTGGCGGATGCTGGCG 1080  
 QY 1081 AAGGCTTTCGACGACGCAAGGCGGATGCTGGCGGATGCTGGCGGATGCTGGCGGATGCTGGCG 1140  
 DB 1081 AAGGCTTTCGACGACGCAAGGCGGATGCTGGCGGATGCTGGCGGATGCTGGCGGATGCTGGCG 1140

Db 1081 AAGGCTTTGACGACGAGAGGGGCTTACTGTCCGAAAGGGTGGCCACGACCTGGGG 1140  
 QY 1141 GGGCGTTACAGGAGCCGCGGCTGCTGGGCTGACTGCTGCGCGCTGTTACCGCAG 1200  
 Db 1141 GTGCGTTACAGGAGCCGCGGCTGCTGGGCTGACTGCTGCGCGCTGTTACCGCAG 1200  
 QY 1201 CMCAGTCG 1209  
 Db 1201 CACACNTCG 1209  
 RESULT 5  
 LOCUS A89746 1235 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 3 from Patent WO9832862.  
 ACCESSION A89746  
 VERSION A89746.1 GI:6738280  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1235)  
 FLOHE, L. and Singh, M.  
 TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
 JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source  
 BASE COUNT 236 a 394 c 382 g 220 t 3 others  
 ORIGIN  
 Query Match 99.8%; Score 1207; DB 5; Length 1235;  
 Best local similarity 99.8%; Pred. No. 1.5e-165;  
 Matches 1207; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 541 GCGGGGTGCCCCGCGTGAACCGCGCGAGCTGTGTATTCGGCGCCGACCGCGC 600  
 Db 541 GGGGGGTGCCCCGCGTGAACCGCGCGAGCTGTGTATTCGGCGCCGACCGCGC 600  
 QY 601 TACAAAGCAGCGCCGATCGCAAGGCGATGGGCGGACCGTTACGGTTTACATCAAC 660  
 Db 601 TACAAAGCAGCGCCGATCGCAAGGCGATGGGCGGACCGTTACGGTTTACATCAAC 660  
 QY 661 A1CGCAAACTTCGCAACTGACGCGGAGTTCTGGCGCGGATCCAGCTGCTACTCA 720  
 Db 661 ATGCAAAACTTCGCAACTGACGCGGAGTTCTGGCGCGGATCCAGCTGCTACTCA 720  
 QY 721 TCGGCTTACGAGCTTCGAGGCTGCGCTCAAGCTGCCAGCTGTGATTTGGGCGCTCTG 780  
 Db 721 TCGGCTTACGAGCTTCGAGGCTGCGCTCAAGCTGCCAGCTGTGATTTGGGCGCTCTG 780  
 QY 781 GTGCGAGCGCGCAAGGCGCCCAATTAGTCTGAATTACTGTGCGCATATGAACCA 840  
 Db 781 GTGCGAGCGCGCAAGGCGCCCAATTAGTCTGAATTACTGTGCGCATATGAACCA 840  
 QY 841 GGTGGGTACTGTGTGATATAGCCATGACAGGCGGCTGTTGAAAGCTCAGACCG 900  
 Db 841 GTGGGCTACTGTGTGATATAGCCATGACAGGCGGCTGTTGAAAGCTCAGACCG 900  
 QY 901 ACCACTACGACACCGCGAGCTTGGCGGTGACAGACAGCGTTTACTGCGTGGCGAAC 960  
 Db 901 ACCACTACGACACCGCGAGCTTGGCGGTGACAGACAGCGTTTACTGCGTGGCGAAC 960  
 QY 961 ATGCCCGCTGCTGCCAGACGCTGACCTACGCTACGCTGACCAACGCGACGATCCGTAT 1020  
 Db 961 ATGCCCGCTGCTGCCAGACGCTGACCTACGCTACGCTGACCAACGCGACGATCCGTAT 1020  
 QY 1021 GTGCTCGAGCTTGGCGACATGCTGGCGGCGGCTGCGCTGGAATCCGGCCTAGCC 1080  
 Db 1021 GTGCTCGAGCTTGGCGACATGCTGGCGGCGGCTGCGCTGGAATCCGGCCTAGCC 1080  
 QY 1081 AAGGCTTTTGCAGCAGCAAGAGGCGCTTACTGTCCGAAGGGTGGCCAGCTGAGGG 1140  
 Db 1081 AAGGCTTTTGCAGCAGCAAGAGGCGCTTACTGTCCGAAGGGTGGCCAGCTGAGGG 1140  
 QY 1141 GTGCGTTACCGAGCCGCGCAGCGTGTGCGCTGACTCTGCGCGCTGTTACGCGAG 1200  
 Db 1141 GTGCGTTACCGAGCCGCGCAGCGTGTGCGCTGACTCTGCGCGCTGTTACGCGAG 1200  
 QY 1201 CMCAGTCG 1209  
 Db 1201 CACACNTCG 1209  
 RESULT 6  
 LOCUS A87608 1228 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 5 from Patent WO9836089.  
 ACCESSION A87608  
 VERSION A87608.1 GI:6736248  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1228)  
 FLOHE, L. and Singh, M.  
 TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source  
 BASE COUNT 236 a 391 c 382 g 219 t  
 ORIGIN



Db 421 CAGACCGCCGAGGCGCACTACCCCTGCTTGGCCCGATGAJCGAAGTCGCGGTGACTC 480  
 QY 481 GCGGCCAGAGTTGGGCTTACCACTGATGCCAAGGAGGCGCGGCTGTGCTGATG 540  
 Db 481 GCGGCCAGAGTTGGGCTTACCACTGATGCCAAGGAGGCGCGGCTGTGCTGATG 540  
 QY 541 GCGGGGCTCCGCGCTCGAAGCGGCGGAGTGTGTATGCGGCGCGGAGCGGCGG 600  
 Db 541 GCGGGGCTCCGCGCTCGAAGCGGCGGAGTGTGTATGCGGCGCGGAGCGGCGG 600  
 QY 601 TACACGACGCGCGCATGCGCAAGGCGGATGGCGGCGGCGGCTTACGCTTACATCAAC 660  
 Db 601 TACACGACGCGCGCATGCGCAAGGCGGATGGCGGCGGCGGCTTACGCTTACATCAAC 660  
 QY 661 ATCGACAACTTCGGCAATCGACGCGGATTTGGGCGCGGATTCACACTGCTACTCA 720  
 Db 661 ATCGACAACTTCGGCAATCGACGCGGATTTGGGCGCGGATTCACACTGCTACTCA 720  
 QY 721 TCGGCTTACGAGCTGAGGCTGCGTCAAACTGCGGAGCTGCTGATTTGGGCGGCTCTG 780  
 Db 721 TCGGCTTACGAGCTGAGGCTGCGTCAAACTGCGGAGCTGCTGATTTGGGCGGCTCTG 780  
 QY 781 GTGCGAGGCGCCAAAGGCGCAAAATTAATAGTCTGCAATTCATTTGTGCGCATATGAACA 840  
 Db 781 GTGCGAGGCGCCAAAGGCGCAAAATTAATAGTCTGCAATTCATTTGTGCGCATATGAACA 840  
 QY 841 GGTGCGGTACTGTGTATATAGCATCGACAGGCGGCTGTTTTCGAAGGCTACACCG 900  
 Db 841 GGTGCGGTACTGTGTATATAGCATCGACAGGCGGCTGTTTTCGAAGGCTACACCG 900  
 QY 901 ACCACCTAGACACCGAGCTTGGCGGTGCGGCTGACAGACGCTGTTTACTGCGGCGGAC 960  
 Db 901 ACCACCTAGACACCGAGCTTGGCGGTGCGGCTGACAGACGCTGTTTACTGCGGCGGAC 960  
 QY 961 ATGCGCGCTCGGTGCGGAGAGACGCTGACCTGACGCTGACCAAGCGAGATGCGGTAT 1020  
 Db 961 ATGCGCGCTCGGTGCGGAGAGACGCTGACCTGACGCTGACCAAGCGAGATGCGGTAT 1020  
 QY 1021 GTGCTGAGCTTCCGACCATGGCTGGGCGGCGGCTGCGGCTGCAATCCGCGACTAGCC 1080  
 Db 1021 GTGCTGAGCTTCCGACCATGGCTGGGCGGCGGCTGCGGCTGCAATCCGCGACTAGCC 1080  
 QY 1081 AAGGCTTTCGACGACGAGAGGCGGCTTACTGTCCGAAGGCGGCGGCGGCGGAG 1140  
 Db 1081 AAGGCTTTCGACGACGAGAGGCGGCTTACTGTCCGAAGGCGGCGGCGGCGGAG 1140  
 QY 1141 GTGCGCTTACCGAGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
 Db 1141 GTGCGCTTACCGAGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
 QY 1201 CACACGCTG 1209  
 Db 1201 CACACGCTG 1209

RESULT 8  
 LOCUS A87607 1237 bp DNA  
 DEFINITION Sequence 4 from Patent WO9836089.  
 ACCESSION A87607  
 VERSION A87607.1 GI:6736247  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1237)  
 AUTHORS Flohe, L. and Singh, M.  
 TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 1. 1237  
 source /organism="unidentified"

BASE COUNT 236 a 394 c 386 g 221 t  
 ORIGIN  
 Query Match 99.8%; Score 1206.4; DB 5; Length 1237;  
 Best Local Similarity 99.8%; Pred. No. 1.9e-165;  
 Matches 1207: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCTTGACAGATTAAATGAACTTCTTCACTGAAAGCTACGATATGAGAGGGTAAATC 60  
 Db 1 ATCTTGACAGATTAAATGAACTTCTTCACTGAAAGCTACGATATGAGAGGGTAAATC 60  
 QY 61 ATGCGCGTGGTATTCGACGAGACCAAAACAAAGAAATTCGGGCGGCGGCGGCGG 120  
 Db 61 ATGCGCGTGGTATTCGACGAGACCAAAACAAAGAAATTCGGGCGGCGGCGGCGG 120  
 QY 121 GCGGCGTTCGCGAACTAACCGTGTGGCCATGAGGTCTCATCCAGGAGGTCCGGA 180  
 Db 121 GCGGCGTTCGCGAACTAACCGTGTGGCCATGAGGTCTCATCCAGGAGGTCCGGA 180  
 QY 181 GAGGCTGCGCTATACCGAGCGGATTTCAAGCGCGACGCGCGCAACTGGTCGGCAC 240  
 Db 181 GAGGCTGCGCTATACCGAGCGGATTTCAAGCGCGACGCGCGCAACTGGTCGGCAC 240  
 QY 241 GCGGACCAAGTGTGGGCGGCGGCGGATTTATGCTCAAGGTCAAGAACCGATACGCGG 300  
 Db 241 GCGGACCAAGTGTGGGCGGCGGCGGATTTATGCTCAAGGTCAAGAACCGATACGCGG 300  
 QY 301 GAATACGCGCGCGCTGCGACGAGCGGAGATCTTTCACGCTTCTGCAATTTGGCGCGTCA 360  
 Db 301 GAATACGCGCGCGCTGCGACGAGCGGAGATCTTTCACGCTTCTGCAATTTGGCGCGTCA 360  
 QY 361 CGTGTTCGACCGATGCGTGTGATTCGCGACAGTCAATTTGCTACGAGACCGTC 420  
 Db 361 CGTGTTCGACCGATGCGTGTGATTCGCGACAGTCAATTTGCTACGAGACCGTC 420  
 QY 421 CAGACCGCGGAGGCGGCACTACCGCTGCTGCGCGGATGAGAGTCCGCGGTGACATC 480  
 Db 421 CAGACCGCGGAGGCGGCACTACCGCTGCTGCGCGGATGAGAGTCCGCGGTGACATC 480  
 QY 481 GCGGCGCAGGTTGGCGCTTACCACTGATGCAACCAAGGCGGCGGCGGTGTGCTGATG 540  
 Db 481 GCGGCGCAGGTTGGCGCTTACCACTGATGCAACCAAGGCGGCGGCGGTGTGCTGATG 540  
 QY 541 GCGGCGGTCGCGGCGTGCACCGGCGGAGCTGTGTGATGCGGCGGCGGCGGCGG 600  
 Db 541 GCGGCGGTCGCGGCGTGCACCGGCGGAGCTGTGTGATGCGGCGGCGGCGGCGG 600  
 QY 601 TACAAAGCGAGCGCGCATGCGCAAAAGCATGGCGGCGGCGGCTTACGATCAAC 660  
 Db 601 TACAAAGCGAGCGCGCATGCGCAAAAGCATGGCGGCGGCGGCTTACGATCAAC 660  
 QY 661 ATCGACAACTTCGCAATCGACGCGGATTCGCGGCGGATCCACACTGCTACTCA 720  
 Db 661 ATCGACAACTTCGCAATCGACGCGGATTCGCGGCGGATCCACACTGCTACTCA 720  
 QY 721 TCGGCTTACGAGCTGAGGCTGCGTCAAACTGCGGAGCTGCTGATTTGGGCGGCTCTG 780  
 Db 721 TCGGCTTACGAGCTGAGGCTGCGTCAAACTGCGGAGCTGCTGATTTGGGCGGCTCTG 780  
 QY 781 GTGCGAGGCGCCAAAGGCGGCGGCAAAATTAATAGTCTGCAATTCATTTGTGCGCATATGAACA 840  
 Db 781 GTGCGAGGCGCCAAAGGCGGCGGCAAAATTAATAGTCTGCAATTCATTTGTGCGCATATGAACA 840  
 QY 841 GGTGCGGTACTGTGTATATAGCATCGACAGGCGGCTGTTTTCGAAGGCTACACCG 900  
 Db 841 GGTGCGGTACTGTGTATATAGCATCGACAGGCGGCTGTTTTCGAAGGCTACACCG 900  
 QY 901 ACCACCTAGACACCGAGCTTGGCGGTGCGGCTGACAGACGCTGTTTACTGCGGCGGAGAC 960  
 Db 901 ACCACCTAGACACCGAGCTTGGCGGTGCGGCTGACAGACGCTGTTTACTGCGGCGGAGAC 960

QY 961 ATGCCCGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGACGCTAT 1020  
 DB 961 ATGCCCGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGACGCTAT 1020  
 QY 1021 GTGCTGACGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 1080  
 DB 1021 GTGCTGACGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 1080  
 QY 1081 AAGGCTGCTGACGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 1140  
 DB 1081 AAGGCTGCTGACGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 1140  
 QY 1141 GTGCTGACGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 1200  
 DB 1141 GTGCTGACGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 1200  
 QY 1201 CACACGCTG 1209  
 DB 1201 CACACGCTG 1209

RESULT 9  
 A89747 1237 bp DNA PAT 22-JAN-2000  
 LOCUS A89747 Sequence 4 from Patent WO9832862.  
 DEFINITION A89747  
 ACCESSION A89747.1 GI:6738281  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 unclassified.  
 1 (bases 1 to 1237)  
 AUTHORS Flohe, L. and Singh, M.  
 TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM  
 JOURNAL Patent: WO 9832862-A 30-JUL-1998  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source location/Qualifiers  
 1..1237  
 /organism="unclassified"  
 /db\_xref="taxon:32844"  
 BASE COUNT 236 a 394 c 386 g 221 t  
 ORIGIN

Query Match 99.8%; Score 1206.4; DB 5; Length 1237;  
 Best Local Similarity 99.8%; Pred. No. 1.9e-165;  
 Matches 1207; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 ATCTGACATTAATGAACTTCTTACACTGAGCTGACGCTGACGAGGCTATATC 60  
 DB 1 ATCTGACATTAATGAACTTCTTACACTGAGCTGACGCTGACGAGGCTATATC 60  
 QY 61 ATGCGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 120  
 DB 61 ATGCGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 120  
 QY 121 GCGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 180  
 DB 121 GCGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 180  
 QY 181 GAGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 240  
 DB 181 GAGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 240  
 QY 241 GCGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 300  
 DB 241 GCGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 300  
 QY 301 GAATACGCGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 360  
 DB 301 GAATACGCGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 360  
 QY 361 CGTGTGACGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 420

DB 361 CGTGTGACGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 420  
 QY 421 CAGACGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 480  
 DB 421 CAGACGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 480  
 QY 481 GCGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 540  
 DB 481 GCGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 540  
 QY 541 GCGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 600  
 DB 541 GCGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 600  
 QY 601 TACACGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 660  
 DB 601 TACACGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 660  
 QY 661 ATGCAAACTTGGCACTGACGCTGACCTGACCTGACCAACGACGCTAT 720  
 DB 661 ATGCAAACTTGGCACTGACGCTGACCTGACCTGACCAACGACGCTAT 720  
 QY 721 TCGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 780  
 DB 721 TCGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 780  
 QY 781 GTGCTGACGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 840  
 DB 781 GTGCTGACGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 840  
 QY 841 GGTGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 900  
 DB 841 GGTGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 900  
 QY 901 ACCACTGACGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 960  
 DB 901 ACCACTGACGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 960  
 QY 961 ATGCGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 1020  
 DB 961 ATGCGGCTGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 1020  
 QY 1021 GTGCTGACGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 1080  
 DB 1021 GTGCTGACGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 1080  
 QY 1081 AAGGCTGCTGACGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 1140  
 DB 1081 AAGGCTGCTGACGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 1140  
 QY 1141 GTGCTGACGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 1200  
 DB 1141 GTGCTGACGCTGCTGCGGAGAGCTGACCTGACGCTGACCAACGACGCTAT 1200  
 QY 1201 CACACGCTG 1209  
 DB 1201 CACACGCTG 1209

RESULT 10  
 MTW002 56414 bp DNA BCT 17-JUN-1998  
 LOCUS MTW002  
 DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.  
 ACCESSION AL008967 AL123456  
 VERSION AL008967.1 GI:3261491  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mycobacterium tuberculosis.  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Corynebacterineae; Mycobacteriaceae;  
 Mycobacterium.  
 REFERENCE 1 (bases 1 to 56414)

**AUTHORS** Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gas S., Barry III C.E., Tekla F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtz S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulten J.E., Taylor K., Whitehead S. and Barrall B.G.

**TITLE** Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence

**JOURNAL** Nature 393 (6685), 537-544 (1998)

**REMARK** Erratum: [[published erratum appears in Nature 1998 Nov 12;396(6707):190]]  
2 (bases 1 to 56414)  
Parkhill, J.

**REFERENCE** Direct Submission  
Submitted (11-JUN-1998) submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
On Jun 27, 1998 this sequence version replaced gi:2624256.

**COMMENT** Notes:  
Details of *M. tuberculosis* sequencing at the Sanger Centre are available on the World Wide Web.  
(URL: [http://www.sanger.ac.uk/Projec\\_s/M.tuberculosis/](http://www.sanger.ac.uk/Projec_s/M.tuberculosis/)) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.  
Gene prediction was based on a Hidden Markov Model of TB genes implemented in TParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.  
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtc, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
Location/Qualifiers  
1..56414  
/organism="Mycobacterium tuberculosis"  
/strain="H37Rv"  
/db\_xref="taxon:1773"  
/note="56414  
/note="fragment designated v002. Does not represent a physical clone"  
1..233  
/organism="Mycobacterium tuberculosis"  
/strain="H37Rv"  
/db\_xref="taxon:1773"  
/clone="Y154"  
/complement(3..527)  
/gene="recX"  
/complement(3..527)  
/gene="recX"  
/note="Rv2736c, (MTV002.02c), len: 174 aa, recX, similar to eg. RECX\_PSEAE P37860 regulatory protein recX from *Pseudomonas aeruginosa* (153 aa), fasta scores: opt: 161 z-score: 257.2 E(): 3.6e-07, 30.7% identity in 137 aa overlap. Overlaps and extends CDS from overlapping cosmid MTCY154.16c"  
/codon\_start=1  
/transl\_table=11  
/product="recX"  
/protein\_id="CA15532.1"  
/db\_xref="GI:2624257"  
/translation="MTVSCPSPSTSRERQAPALICRLTARSRAELAGOLAKRGYPEDIGRVLDRALAVGLVDITPAEDWVOSRRANKSKRALAEIHLAKGVDDVITTVGGIDAGAEGRGAERLVARLRREVLIIDGDEARVSRVLVAMLARRGYGTACCEV VIAELAAERERRRV"

**gene** complement(493..2865)  
/gene="recX"  
/complement(493..2865)  
/note="Rv2737c, (MTV002.02c), len: 790 aa, recX, identical to RECA\_MYCRU p26345 recA protein (790 aa). Contains self-splicing protein element (intron) from 2294 to 969 (c) similar to intron II from TR:E332317 (EMBL:Y13030) DNA-directed DNA polymerase (EC 2.7.7.7) from *Thermococcus* sp. (1829 aa), fasta scores: opt: 81 z-score: 235.2 E(): 6e-06, 24.6% identity in 183 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop), PS00321 recA signature, and PS00881 protein splicing signature. See Davis et al, (1992) Cell 71(2):201-210"  
/codon\_start=1  
/transl\_table=11  
/product="recA"  
/protein\_id="CA15533.1"  
/db\_xref="GI:2624258"  
/db\_xref="SWISS-PROT:P26345"  
/translation="MTCTPDREKALELAVNAIEKSGKGYMRUGDEARQIPSYIPGSIADVAIGIGLPRGVETIEGPSSGKTVLHAHVAHQAGVAFDAEALADP DYARKLGVDTSLVSPDGEQALADMLIRGALDIVVDAALVPALEEGM GDSHVGQARLMSQALRMTGALNNGSTTAIFINQDRKIGVMGSEETGVALNF YASVMDVRYEITLKDGNAVGNRTRVVYVNRKCLAGTRIFDPVGTTHIEDVVG RKP1HVVAADKGLTHARPVSWFDGTRVDVIGRIAGATVATPDHKTLYEGWRA AGELRKGDVAPRRFGSDAP1PDHARLGLIGDGDGVGKTPINFTNVR ALIDVTRVANTLGCARHPGRTSLAHHRGKNGVADLCQAGITKLAETITRN WFFEDVADIVGNLIFGLFESDWSVSEQFARLVGTTTSEQLAHIMHLRLEF GSTVADYDPTOKRPSIVNGRIQSRQVEFARISGMNVYAFASVPMGRNALIQ AIPEXTQGRSGSATYLAEMTDAVLVLDERTVTOEALAMGVASGDPGCMKV LAGSLRRDRQALADALDKFLDMLAEELRYVIREVPTRRARFEDLEELHTL VAEVYVNCSPSPKQAEFLYKKGISREGLSDMKVDGLIRKSCAMPTVEEQIG OGKERNRFLVENDVADDEIKRKIKELIGIAVVTDDPSNDGVLPAVDF"  
complement(790..2112)  
/gene="recA"  
/note="recA intein"  
complement(790..807)  
/gene="recA"  
/note="PS00881 Protein splicing signature"  
complement(2197..2223)  
/gene="recA"  
/note="PS00321 recA signature"  
complement(2644..2667)  
/gene="recA"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
complement(2873..2877)  
/note="possible RBS upstream of Rv2737c"  
complement(3247..3453)  
/gene="Rv2738c"  
complement(3247..3453)  
/gene="Rv2738c"  
/note="Rv2738c, (MTV002.03c), len: 68 aa, Unknown, N-terminus is highly similar to the N-terminus of the upstream ORF MTV002.07c (78.4% identity in 37 aa overlap); also similar to A10209581SC448.5 Streptomyces coelicolor cosmid 4H8 (64aa) opt: 185 z-score: 283.5 E(): 2.9e-08; 39.7% identity in 63 aa overlap.  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein Rv2738c"  
/protein\_id="CA15534.1"  
/db\_xref="GI:2624260"  
/db\_xref="SPTREMBL:O33281"  
/translation="MLAGVRLTEFERHAFHGAAYGSSVLLDHYLTGFDGRSAQAIEDGVEPRDVMRALCADFVPHDRW"  
complement(3464..4630)  
/gene="Rv2739c"  
complement(3464..4630)  
/gene="Rv2739c"  
/note="Rv2739c, (MTV002.04c), len: 388 aa, Probable transferase, similar to eg. TR:O51560 (EMBL:L28170) *Pseudomonas aeruginosa* 'chamosyl' transferase (426 aa),



fasta\_scores: opt: 178 z-score: 226.3 E(): 1.9e-05, 25.9% identity in 425 aaooverlap. Equivalent to Mycobacterium leprae protein MLCB3.02c (392 aa); fasta scores 9p[294723]MLCB3\_2 Mycobacterium leprae cosmid B33 opt: 2112 z-score: 2364.5 E(): 0; 80.9% identity in 388 aa overlap

overlapp  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein Rv2739c"  
/protein\_id="CAI15535.1"  
/db\_xref="GI:2624261"  
/db\_xref="SPTREMBL:O33282"  
/translation="MRVAVVAGPDPGSHSPALALCORPRAADPTLTGYEMLEAAR  
AAGIDAVELDLAATRDLDAGARIHRAQAQAVNVPRALPELAVSVATACG  
MAELGLIPWEINLPHPLPLPSKGLPISGAGTGRRLRDLATMALGSRMRG  
LROBAVREVEIGLPARDGRLRLIATLPALEVPDPMPAAVAVVGPAPHEPTRYLA  
IPAGTGVVAVVAPSTALGTAGLTVLALQSLTPTGTVSGSRVLSAGDLTPPA  
AVAGLSQAEHLTRADLVICGGHGMVAKTLIAGPVPVAVVQSRVQSLIARVVOGS  
AVLIRPLADALVAANVEVLSPPREARRAASVAGAADPVRVCHDALALAG"  
4674.5123  
/gene="Rv2740"  
4674.5123  
/note="Rv2740, (MTV002.05), len: 149 aa, unknown, similar  
to M. leprae hypothetical protein TR\_049850 (EMBL:294723)  
MLCB3.03 (B223\_F3.140) (178 aa), fasta scores: opt: 498  
z-score: 538.9 E(): 7.3e-23, 51.6% identity in 161 aa  
overlap"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein Rv2740"  
/protein\_id="CAI1536.1"  
/db\_xref="GI:2624262"  
/db\_xref="SPTREMBL:O33283"  
/translation="MAELTSPETPEAIRAVEAFNLALONEDPDTVAIGDDL  
VYENGFSRIRGRRATLRLRMGCRVFEYKIRHIGDGAALTEPRTDALIIGPLRV  
QPNVCGVEVDGRTTLRDRYFDYDMKGLKGLVALVPSLAKTL"  
5355.6932  
/gene="PE\_PGRS"  
5355.6932  
/note="Rv2741, (MTV002.06), len: 525 aa; Member of M.  
tuberculosis PE\_PGRS subfamily, similar to many eg.

Query Match 99.8%; Score 1206.4; DB 1; Length 56414;  
Best Local Similarity 99.8%; Pred. No. 8.1e-166;  
Matches 1207; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

gene  
CDS  
1 ANCTTGCAATTAATCGAATCTTTCTTCACTGAAAGCGTACAGTATCGAGAGGGTAAATC 60  
38201 ATCTTGCAATTAATCGAATCTTTCTTCACTGAAAGCGTACAGTATCGAGAGGGTAAATC 38260  
61 ATGCGCGTGGTATTCGACGACGACAAACAAACGAATTCGGGGTGGCCATCAACCCG 120  
38261 ATGCGCGTGGTATTCGACGACGACAAACAAACGAATTCGGGGTGGCCATCAACCCG 38320  
121 GCCGGCGTGGCGAACTAACCCGCTGGTGGCATGAGGTGCTATCCAGGACAGTCCCGGA 180  
38321 GCCGGCGTGGCGAACTAACCCGCTGGTGGCATGAGGTGCTATCCAGGACAGTCCCGGA 38380  
181 GAGGGCTGGCTATCAGCAGCGAGGATTCAGGGCGGCGGCGCAACTGGTGGCGAAC 240  
38381 GAGGGCTGGCTATCAGCAGCGAGGATTCAGGGCGGCGGCGCAACTGGTGGCGAAC 38440  
241 GCCGACAGAGTGGGCGCAGCTGATTATTGCTCAACJTCGAAGAAACCGATAGCGGCG 300  
38441 GCCGACAGAGTGGGCGCAGCTGATTATTGCTCAACJTCGAAGAAACCGATAGCGGCG 38500  
301 GAATACGGCGCGCTGCGACAGCGGAGATCTTGTACGTTTGTGCAATTTGGCCCGCTCA 360  
38501 GAATACGGCGCGCTGCGACAGCGGAGATCTTGTACGTTTGTGCAATTTGGCCCGCTCA 38560  
361 CGTGCTGACAGATGCGTGTGTGATTCGGGACACAGCTCAATTTGCCATACAGACCGTC 420

Db 38561 CGTCTTACACAGATGCGTTTGTGATTCGGGACCCACGATCAATTTGCTTACAGAACCTTC 38620  
QY 421 CAGACCGCGCAGCGGCGCACTACCCCTGCTTCCCGCATGAGCGAAGTGGCGGTGACTC 480  
Db 38621 CAGACCGCGCAGCGGCGCACTACCCCTGCTTCCCGCATGAGCGAAGTGGCGGTGACTC 38680  
QY 481 GCCGCGCAGGTTGGCGCTTACCACCTGATGGAACCCAAAGGGGCGCGGTGTCTGATG 540  
Db 38681 GCCGCGCAGGTTGGCGCTTACCACCTGATGGAACCCAAAGGGGCGCGGTGTCTGATG 38740  
QY 541 GCGGGGTGGCGCGCTGCAACCGCGGAGCTCGTGTGATCGCGCGCGGACCGCGCG 600  
Db 38741 GCGGGGTGGCGCGCTGCAACCGCGGAGCTCGTGTGATCGCGCGCGGACCGCGCG 38800  
QY 601 TACAACGAGCGCGCATGCGCAACGCGATGGGCGCGACCGTTACGTTTACATCAAC 660  
Db 38801 TACAACGAGCGCGCATGCGCAACGCGATGGGCGCGACCGTTACGTTTACATCAAC 38860  
QY 661 ATGACAAACTTGGCAACTGACGCGGAGTTCGGGCGGATCCACATCGCTACTCA 720  
Db 38861 ATGACAAACTTGGCAACTGACGCGGAGTTCGGGCGGATCCACATCGCTACTCA 38920  
QY 721 TCGGCTTACGAGCTGAGAGGTGCGGTCAACGTCGACCTGTGATTTGGGCGCGTCTG 780  
Db 38921 TCGGCTTACGAGCTGAGAGGTGCGGTCAACGTCGACCTGTGATTTGGGCGCGTCTG 38980  
QY 781 GTGCGAGCGCGCAAGCAGCAACCAATATGCTCGCAATTCATCTTGGCGCATATGAACCA 840  
Db 38981 GTGCGAGCGCGCAAGCAGCAACCAATATGCTCGCAATTCATCTTGGCGCATATGAACCA 39040  
QY 841 GTGCGAGTACTGTGATATATAGCATGACGACGAGGCGGTGTTGGAAGGTCAAGACG 900  
Db 39041 GTGCGAGTACTGTGATATATAGCATGACGACGAGGCGGTGTTGGAAGGTCAAGACG 39100  
QY 901 ACCACTTACGACCAACCGCAGCTGCGGTGACGACGACGCTGTTTACTGCTGGCGAAC 960  
Db 39101 ACCACTTACGACCAACCGCAGCTGCGGTGACGACGACGCTGTTTACTGCTGGCGAAC 39160  
QY 961 ATGCGCGCGTGGTGGCGAAGACGCTGACCTACGCGGTGACCAACGGGAGATGCCGTAT 1020  
Db 39161 ATGCGCGCGTGGTGGCGAAGACGCTGACCTACGCGGTGACCAACGGGAGATGCCGTAT 39220  
QY 1021 GTGCTCGAGTGTGCGACCATGCGTGGCGGCGGCGGTGATCCGGCATAGCC 1080  
Db 39221 GTGCTCGAGTGTGCGACCATGCGTGGCGGCGGCGGTGATCCGGCATAGCC 39280  
QY 1081 AAAGCTTTTGGACGACGAAAGCGCGCTTACTGTCCGAACGGGTGGCCACGACTGGGG 1140  
Db 39281 AAAGCTTTTGGACGACGAAAGCGCGCTTACTGTCCGAACGGGTGGCCACGACTGGGG 39340  
QY 1141 GTGCGTTCACGAGCGCGGCTTACGCTGCGCTGCGCTGCGCGCTGCTTACGCGGAG 1200  
Db 39341 GTGCGTTCACGAGCGCGGCTTACGCTGCGCTGCGCTGCGCGCTGCTTACGCGGAG 39400  
QY 1201 CNCAAGTGG 1209  
Db 39401 CACAGTGG 39409

RESULT 11  
A89752  
LOCUS A89752 1208 bp DNA  
DEFINITION Sequence 9 from Patent W09832862.  
ACCESSION A89752  
VERSION A89752.1 GI:6738285  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1208)  
AUTHORS Flohe, L. and Singh, M.  
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM



JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source 1. .1208  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 228 a 391 c 373 g 216 t  
 ORIGIN

Query Match 99.8%; Score 1201; DB 5; Length 1208;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-164;  
 Matches 1201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCTGCAAGATTATCGAAGCTTTCTTCACTGAAGCGTACAGTATCGAGAGGGTAATC 60  
 DB 1 ATCTGCAAGATTATCGAAGCTTTCTTCACTGAAGCGTACAGTATCGAGAGGGTAATC 60

QY 61 ATGCGCGTGGTATTCGACGAGACCAAAACAAGAAATCCGGGTGGCCATCAACCCCG 120  
 DB 61 ATGCGCGTGGTATTCGACGAGACCAAAACAAGAAATCCGGGTGGCCATCAACCCCG 120

QY 121 GCCGCGCTGCGGGAACCTAACCCGCTGTGGCCATGAGTGTCTCATCCAGGCAAGTCCGGA 180  
 DB 121 GCCGCGCTGCGGGAACCTAACCCGCTGTGGCCATGAGTGTCTCATCCAGGCAAGTCCGGA 180

QY 181 GAGGCTGCGCTATCACCGACGCGGATTTCAAAGCGGCGCAACTGGTGGCACC 240  
 DB 181 GAGGCTGCGCTATCACCGACGCGGATTTCAAAGCGGCGCAACTGGTGGCACC 240

QY 241 GCCGACAGGTGTGGGCGGAGCGGCTATTTATGCTCAAGGTCAAAGAACCGATACGGCG 300  
 DB 241 GCCGACAGGTGTGGGCGGAGCGGCTATTTATGCTCAAGGTCAAAGAACCGATACGGCG 300

QY 301 GAATACGGCCCGCTCGACACGCGGAGATCTTGTACAGTTCTTGCATTGGCCCGCTCA 360  
 DB 301 GAATACGGCCCGCTCGACACGCGGAGATCTTGTACAGTTCTTGCATTGGCCCGCTCA 360

QY 361 CGTGGTTCACCGATGCGTTGTTGATTCGCGGACACGTCATTTGCCAGACCGTTC 420  
 DB 361 CGTGGTTCACCGATGCGTTGTTGATTCGCGGACACGTCATTTGCCAGACCGTTC 420

QY 421 CAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480  
 DB 421 CAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480

QY 481 GCCGCGCGGCTTGGCGCTTACCACTGATGCGAACCAAGGCGGCGGCGGCTGCTGATG 540  
 DB 481 GCCGCGCGGCTTGGCGCTTACCACTGATGCGAACCAAGGCGGCGGCGGCTGCTGATG 540

QY 541 GCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
 DB 541 GCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600

QY 601 TACAACGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660  
 DB 601 TACAACGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660

QY 661 ATGCAACAATTGCGCAACTGCGACCGCGGATTCGCGGCGGCGGCGGCGGCGGCGG 720  
 DB 661 ATGCAACAATTGCGCAACTGCGACCGCGGATTCGCGGCGGCGGCGGCGGCGGCGG 720

QY 721 TCGGCGCTACGAGCTGAGGCGTCCGTAAGTGGCGGCGGCGGCGGCGGCGGCGGCGG 780  
 DB 721 TCGGCGCTACGAGCTGAGGCGTCCGTAAGTGGCGGCGGCGGCGGCGGCGGCGGCGG 780

QY 781 GTGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840  
 DB 781 GTGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840

QY 841 GGTGGGCTACTGTTGGATATACCATCGACGAGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
 DB 841 GGTGGGCTACTGTTGGATATACCATCGACGAGGCGGCGGCGGCGGCGGCGGCGGCGG 900

QY 901 ACCACTACGACACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960  
 DB 901 ACCACTACGACACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960

QY 961 ATGCGCGCTGGTGGCGGAGACGCTTACGCGGCTGACCAACGCGGAGATCCGCTAT 1020  
 DB 961 ATGCGCGCTGGTGGCGGAGACGCTTACGCGGCTGACCAACGCGGAGATCCGCTAT 1020

QY 1021 GTGCTGAGCTTGGCGGACGATGCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080  
 DB 1021 GTGCTGAGCTTGGCGGACGATGCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080

QY 1081 AAAGCTTTTCGACGACGAGGCGGCTTACTGTCCGAACGGGTGGCCACGACCTGGGG 1140  
 DB 1081 AAAGCTTTTCGACGACGAGGCGGCTTACTGTCCGAACGGGTGGCCACGACCTGGGG 1140

QY 1141 GTGCGCTTACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
 DB 1141 GTGCGCTTACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200

QY 1201 C 1201  
 DB 1201 C 1201

RESULT 12  
 A87609 1235 bp DNA PAT 22-JAN-2000  
 LOCUS Sequence 6 from Patent WO9836089.  
 ACCESSION A87609  
 VERSION A87609.1 GI:6736249  
 KEYWORDS  
 SOURCE un identified.  
 ORGANISM unclassified.

REFERENCE 1 (bases 1 to 1235)  
 Fliche, L. and Singh, M.  
 TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)

FEATURES  
 source 1. .1235  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 235 a 395 c 384 g 220 t 1 others  
 ORIGIN

Query Match 98.8%; Score 1195; DB 5; Length 1235;  
 Best Local Similarity 99.8%; Pred. No. 8.3e-164;  
 Matches 1206; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 ATCTGCAAGATTATCGAAGCTTTCTTCACTGAAGCGTACAGTATCGAGAGGGTAATC 60  
 DB 1 ATCTGCAAGATTATCGAAGCTTTCTTCACTGAAGCGTACAGTATCGAGAGGGTAATC 60

QY 61 ATGCGCGTGGTATTCGACGAGACCAAAACAAGAAATCCGGGTGGCCATCAACCCCG 120  
 DB 61 ATGCGCGTGGTATTCGACGAGACCAAAACAAGAAATCCGGGTGGCCATCAACCCCG 120

QY 121 GCCGCGCTGCGGGAACCTAACCCGCTGTGGCCATGAGTGTCTCATCCAGGCAAGTCCGGA 180  
 DB 121 GCCGCGCTGCGGGAACCTAACCCGCTGTGGCCATGAGTGTCTCATCCAGGCAAGTCCGGA 180

QY 181 GAGGCTGCGCTATCACCGACGCGGATTTCAAAGCGGCGGCGGCGGCGGCGGCGGCGG 240  
 DB 181 GAGGCTGCGCTATCACCGACGCGGATTTCAAAGCGGCGGCGGCGGCGGCGGCGGCGG 240

QY 241 GCCGACAGGTGTGGGCGGAGCGGCTATTTATGCTCAAGGTCAAAGAACCGATACGGCG 300  
 DB 241 GCCGACAGGTGTGGGCGGAGCGGCTATTTATGCTCAAGGTCAAAGAACCGATACGGCG 300

QY 301 GAATACGGCCGCTGCGACAGGGGAGATCTTGTACAGTTCTTGCATTGGCCGCTCA 360  
 DB 301 GAATACGGCCGCTGCGACAGGGG- GATCTTGTACAGTTCTTGCATTGGCCGCTCA 359  
 QY 361 CGTGTTCACAGATGCGTGTGTGATCCGCGACACGATGAATGGCTACGAGACCGTC 420  
 DB 361 CGTGTTCACAGATGCGTGTGTGATCCGCGACACGATGAATGGCTACGAGACCGTC 419  
 QY 421 CAGACCCGCGACGCGACGATCCCTGCTTGTCCCGGAGACGAAAGTCGCGCGTCACTC 480  
 DB 421 CAGACCCGCGACGCGACGATCCCTGCTTGTCCCGGAGACGAAAGTCGCGCGTCACTC 479  
 QY 481 GCGCGCCAGTGGGCGTGTACACCTGATGCGAAACCAAGGGGCGCGGTGTGTATG 540  
 DB 481 GCGCGCCAGTGGGCGTGTACACCTGATGCGAAACCAAGGGGCGCGGTGTGTATG 539  
 QY 541 GCGCGGCTGCGCGGCTGGAACCGCGGACGCTGTGTGATCGCGCGCGCGCGCGG 600  
 DB 541 GCGCGGCTGCGCGGCTGGAACCGCGGACGCTGTGTGATCGCGCGCGCGCGCGG 599  
 QY 601 TACAACGAGCGCGCATGCGCAACGCGATGGCGCGACGCTTACGCTTCTAGACATCAAC 660  
 DB 601 TACAACGAGCGCGCATGCGCAACGCGATGGCGCGACGCTTACGCTTCTAGACATCAAC 659  
 QY 661 ATGCAACAATCTGCGCAACTGAGCGCGGATTTCTGCGCGCGGATCGACATCGCTACTCA 720  
 DB 661 ATGCAACAATCTGCGCAACTGAGCGCGGATTTCTGCGCGCGGATCGACATCGCTACTCA 719  
 QY 721 TCGGCTACGAGCTGAGGAGTGCCTCAAAAGTGCAGACGCTGTGATGGGCGCTCTG 780  
 DB 721 TCGGCTACGAGCTGAGGAGTGCCTCAAAAGTGCAGACGCTGTGATGGGCGCTCTG 779  
 QY 781 GTGCGAGCGCGCAAGGCAACCAATTAATGCTTCGAATTCATCTTGCAGCAT 840  
 DB 781 GTGCGAGCGCGCAAGGCAACCAATTAATGCTTCGAATTCATCTTGCAGCAT 839  
 QY 841 GGTGGGCTACGCTGATATAGCATACGAGCGCGGCTGTGTAAGGCTCAACGACG 900  
 DB 841 GGTGGGCTACGCTGATATAGCATACGAGCGCGGCTGTGTAAGGCTCAACGACG 899  
 QY 901 ACCACCTACGACGACGCGACGCTTCCGCTGCGACGACGCTGTTTACTGCGTGGCGAC 960  
 DB 901 ACCACCTACGACGACGCGACGCTTCCGCTGCGACGACGCTGTTTACTGCGTGGCGAC 959  
 QY 961 ATGCGCGCTCGGTGCGCAAGACGCTGACCTGAGCGCTGACCAAGCGACGATGCGTAT 1020  
 DB 961 ATGCGCGCTCGGTGCGCAAGACGCTGACCTGAGCGCTGACCAAGCGACGATGCGTAT 1019  
 QY 1021 GTGCTGACGCTGCGGACGATGGCTGCGGCGGCGGCTGCGGCTGCAATCCGCACTACCC 1080  
 DB 1021 GTGCTGACGCTGCGGACGATGGCTGCGGCGGCGGCTGCGGCTGCAATCCGCACTACCC 1079  
 QY 1081 AAAGCTCTTTCAGACGACGAGGGGCGTACTGTCCGAAAGGGGTGGCCACGACCTGGGG 1140  
 DB 1081 AAAGCTCTTTCAGACGACGAGGGGCGTACTGTCCGAAAGGGGTGGCCACGACCTGGGG 1139  
 QY 1141 GTGCGCTTACGAGGCGCGGCGGCGGCTGAGCTGAGCTGAGCGCTGAGCGCGGAG 1200  
 DB 1141 GTGCGCTTACGAGGCGCGGCGGCGGCTGAGCTGAGCTGAGCGCGCTGAGCGCGGAG 1199  
 QY 1201 CACACGCTG 1209  
 DB 1201 CACACGCTG 1208

## RESULT 13

LOCUS A89749 1235 bp DNA PAT 22-JAN-2000

DEFINITION Sequence 6 from Patent WO9832862.

ACCESSION A89749

VERSION A89749.1 GI:6738283

KEYWORDS

SOURCE unidentified.

ORGANISM unidentified  
 unclassified.  
 REFERENCE 1 (bases 1 to 1235)  
 AUTHORS Fione, L. and Singh, M.  
 TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM PARINUM  
 JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source 1, 1235  
 location/Qualifiers  
 BASE COUNT 235 a 395 c 384 g 220 t 1 others  
 ORIGIN

Query Match 98.8%; Score 1195; DB 5; Length 1235;  
 Best Local Similarity 99.8%; Pred. No. 8.3e-164;  
 Matches 1206; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 ATCTGCAGATTAAATGAACTTCTTACACCTGAAAGCTGACGATGAGAGGGGTAAATC 60  
 DB 1 ATCTGCAGATTAAATGAACTTCTTACACCTGAAAGCTGACGATGAGAGGGGTAAATC 60  
 QY 61 ATGCGGCTGCTTATTCGACCGGAGCAACCAACGAAATTCGGGTGGCCATCAACCCCG 120  
 DB 61 ATGCGGCTGCTTATTCGACCGGAGCAACCAACGAAATTCGGGTGGCCATCAACCCCG 120  
 QY 121 GCGGCGTCCGGAATCAACCCGCTGCGCATGAGTGTCTCATCCAGGAGTGCCTGGA 180  
 DB 121 GCGGCGTCCGGAATCAACCCGCTGCGCATGAGTGTCTCATCCAGGAGTGCCTGGA 180  
 QY 181 GAGGCTCGGCTATCACCGGAGGATTTCAAGGCGGAGGCGGCGGAGTGTGCGGAC 240  
 DB 181 GAGGCTCGGCTATCACCGGAGGATTTCAAGGCGGAGGCGGAGTGTGCGGAC 240  
 QY 241 GCGGAGGCTGTGGCGCGGCGGCTGATTTATTCGTAAGGCTCAAAAGCGATAGCGGCG 300  
 DB 241 GCGGAGGCTGTGGCGCGGCGGCTGATTTATTCGTAAGGCTCAAAAGCGATAGCGGCG 300  
 QY 301 GAATACGGCCGCTGCGACACGCGGAGATCTGTACGTTCTTGCATTTGGCCGCTCA 360  
 DB 301 GAATACGGCCGCTGCGACACGCGGAGATCTGTACGTTCTTGCATTTGGCCGCTCA 359  
 QY 361 CGTGTTCACAGATGCGTGTGTGATCCGCGACACGATGCGTATGGCGCGGCTCA 420  
 DB 361 CGTGTTCACAGATGCGTGTGTGATCCGCGACACGATGCGTATGGCGCGGCTCA 419  
 QY 421 CAGACCCGCGACGCGACGCTTCCGCTGCGGATGAGCGAAGTCCGCGTGCATC 480  
 DB 421 CAGACCCGCGACGCGACGCTTCCGCTGCGGATGAGCGAAGTCCGCGTGCATC 479  
 QY 481 GCGCGCCAGTGGGCGCTTACACCTGATGCGAAACCAAGGGGCGCGGTGTGTATG 540  
 DB 481 GCGCGCCAGTGGGCGCTTACACCTGATGCGAAACCAAGGGGCGCGGTGTGTATG 539  
 QY 541 GCGGCGTGGCGGCTGCAACCGGCGGAGCTGTGTGATTCGCGCGCGGCGCGGCG 600  
 DB 541 GCGGCGTGGCGGCTGCAACCGGCGGAGCTGTGTGATTCGCGCGCGGCGCGGCG 599  
 QY 601 TACAACGAGCGCGCATGCGCAACGCGATGGCGCGACGCTTACGCTTCTAGACATCAAC 660  
 DB 601 TACAACGAGCGCGCATGCGCAACGCGATGGCGCGACGCTTACGCTTCTAGACATCAAC 659  
 QY 661 ATGCAACAATCTGCGCAACTGAGCGCGGATTTCTGCGCGCGGATCGACATCGCTACTCA 720  
 DB 661 ATGCAACAATCTGCGCAACTGAGCGCGGATTTCTGCGCGCGGATCGACATCGCTACTCA 719  
 QY 721 TCGGCTACGAGCTGAGGAGTGCCTCAAAAGTGCAGACGCTGTGATGGGCGCTCTG 780  
 DB 721 TCGGCTACGAGCTGAGGAGTGCCTCAAAAGTGCAGACGCTGTGATGGGCGCTCTG 779  
 QY 781 GTGCGAGCGCGCAAGGCAACCAATTAATGCTTCGAATTCATCTTGCAGCAT 840  
 DB 781 GTGCGAGCGCGCAAGGCAACCAATTAATGCTTCGAATTCATCTTGCAGCAT 840

Db 780 GNGCAGGCGCCAGGCAACCCAAATTAATCTGATCTGTCGCAATATGAACCA 839  
 QY 841 GGTGGGCTACTGTGGATATAGCATCGACGAGGCGCTGTTGGAAGGCTCAGACCG 900  
 Db 840 GGTGGGCTACTGTGGATATAGCATCGACGAGGCGCTGTTGGAAGGCTCAGACCG 899  
 QY 901 ACCACCTAGACACCCGACGCTGCGGTCAGACGACGCTTTTACTGCGTGCGAAC 960  
 Db 900 ACCACCTAGACACCCGACGCTGCGGTCAGACGACGCTTTTACTGCGTGCGAAC 959  
 QY 961 ATGCCCGCTCGGTGCGGAGAGAGCTGACCTAGCGGCTGACCAACGCGAGATGCCGAT 1020  
 Db 960 ATGCCCGCTCGGTGCGGAGAGAGCTGACCTAGCGGCTGACCAACGCGAGATGCCGAT 1019  
 QY 1021 GTGCTCGAGCTTGGCGGACATGCTGCGGCGGCGGCTGCGGCTGCAATTCGGCACTAGCC 1080  
 Db 1020 GTGCTCGAGCTTGGCGGACATGCTGCGGCGGCGGCTGCGGCTGCAATTCGGCACTAGCC 1079  
 QY 1081 AAAGGCTTTGACGACGACGAGGCGGCTTACTGTCGGAACGCGGCGGCGGCGGCGG 1140  
 Db 1080 AAAGGCTTTGACGACGACGAGGCGGCTTACTGTCGGAACGCGGCGGCGGCGGCGG 1139  
 QY 1141 GTGCGCTTACGACGACGCGGCGGCGGCTGCTGCGGCTGCTGCGGCGGCGGCGGCGG 1200  
 Db 1140 GTGCGCTTACGACGACGCGGCGGCGGCTGCTGCGGCTGCTGCGGCGGCGGCGGCGG 1199  
 QY 1201 CMCAGCTCG 1209  
 Db 1200 CACACGCTCG 1208

## RESULT 14

LOCUS A87611 1235 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 8 from Patent WO9836089.  
 ACCESSION A87611  
 VERSION A87611.1 GI:6736251

## KEYWORDS

SOURCE unidentified.  
 ORGANISM unidentified.  
 unclassified.

REFERENCE 1 (bases 1 to 1235)  
 FLOHE, L. and Singh, M.  
 TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)

## FEATURES

source 1..1235  
 Location/Qualifiers  
 1..1235  
 /organism="unidentified"  
 /db\_xref="taxon:32644"

BASE COUNT 236 a 394 c 385 g 220 t  
 ORIGIN

Query Match 98.8%; Score 1194.4; DB 5; Length 1235;  
 Best Local Similarity 99.8%; Pred. No. 1e-163;  
 Matches 1206; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 ATCTTCAGATTATGAACTTTCTTCACTGACGCTACAGTATCGAGAGGGTAAATC 60  
 Db 1 ATCTTCAGATTATGAACTTTCTTCACTGACGCTACAGTATCGAGAGGGTAAATC 60  
 QY 61 ATGCGGCTGGGTATTCGACCGGAGACCAAAACAAACGA TTCGGGTGACATC..GCCG 120  
 Db 61 ATGCGGCTGGGTATTCGACCGGAGACCAAAACAAACGAATTCGGGTGACATCACCCG 120  
 QY 121 GCGGCGGTGCGGAACTAACCCGCTGTGGCATGAGTGTCTATCCAGGCAAGTGCAGGA 180  
 Db 121 GCGGCGGTGCGGAACTAACCCGCTGTGGCATGAGTGTCTATCCAGGCAAGTGCAGGA 180  
 QY 181 GAGGCTCGGCTATCAGCGAGCGGATTTCAAGCGGCGGCGGCAACTGTCGCGCAC 240  
 Db 181 GAGGCTCGGCTATCAGCGAGCGGATTTCAAGCGGCGGCGGCAACTGTCGCGCAC 240

QY 241 GCGGACGAGTGTGGGCGGACGCTGATTATGCTCAAGTCAAGCAAGACGATAGCGCG 300  
 Db 241 GCGGACGAGTGTGGGCGGACGCTGATTATGCTCAAGTCAAGCAAGACGATAGCGCG 300  
 QY 301 GAATACGCGCGCTGACACAGCGGAGATCTTGTCAAGTCTTGTGCGCGGCTCA 360  
 Db 301 GAATACGCGCGCTGACACAGCGGAGATCTTGTCAAGTCTTGTGCGCGGCTCA 359  
 QY 361 GGTCTTTCACCGATGCGTGTGTGATTCGGGACCAAGTCAATTTGCTACGAGACGTC 420  
 Db 360 GGTCTTTCACCGATGCGTGTGTGATTCGGGACCAAGTCAATTTGCTACGAGACGTC 419  
 QY 421 CAGACCGCGGACGACACTACCCCTGCTTCCCGGATGAGCCGAAGTCCGCGTCACTC 480  
 Db 420 CAGACCGCGGACGACACTACCCCTGCTTCCCGGATGAGCCGAAGTCCGCGTCACTC 479  
 QY 481 GCGGCGGCTTGGCGCTTACACCTGATGGAACCCAAAGGCGGCGGCTGCTGATG 540  
 Db 480 GCGGCGGCTTGGCGCTTACACCTGATGGAACCCAAAGGCGGCGGCTGCTGATG 539  
 QY 541 GCGGCGGCTGCGGCGCTGGAACCGGCGGAGCTGCTGATGCGGCGGCGGCGGCGG 600  
 Db 540 GCGGCGGCTGCGGCGCTGGAACCGGCGGAGCTGCTGATGCGGCGGCGGCGGCGG 599  
 QY 601 TACAACGCGGCGGATGCGCAAGCGCATGGGCGGCGGACGCTTACGCTTCAAGCATCAAC 660  
 Db 600 TACAACGCGGCGGATGCGCAAGCGCATGGGCGGCGGACGCTTACGCTTCAAGCATCAAC 659  
 QY 661 ATGCAAACTTGGGCAACTGAGCGGAGTTCGCGGCGGAGTCCACACTGCTACTCA 720  
 Db 660 ATGCAAACTTGGGCAACTGAGCGGAGTTCGCGGCGGAGTTCACACTGCTACTCA 719  
 QY 721 TCGGCTTACGAGCTGAGAGGCTGCGGCTCAACGTCGCGGCTGATGGGCGGCTCTG 780  
 Db 720 TCGGCTTACGAGCTGAGAGGCTGCGGCTCAACGTCGCGGCTGATGGGCGGCTCTG 779  
 QY 781 GTGCGGCGGCGGACGCGGCAAAATTAATGCTGCAATTAATGCTGCGGCAATGAACCA 840  
 Db 780 GTGCGGCGGCGGACGCGGCAAAATTAATGCTGCAATTAATGCTGCGGCAATGAACCA 839  
 QY 841 GGTGGGCTACTGTGGATATAGCATCGACGAGGCGGCTGTTGGAAGGCTCAGACCG 900  
 Db 840 GGTGGGCTACTGTGGATATAGCATCGACGAGGCGGCTGTTGGAAGGCTCAGACCG 899  
 QY 901 ACCACCTAGACACCCGACGCTGCGGTCAGACGACGCTGTTTACTGCGTGCGAAC 960  
 Db 900 ACCACCTAGACACCCGACGCTGCGGTCAGACGACGCTGTTTACTGCGTGCGAAC 959  
 QY 961 ATGCCCGCTCGGTGCGGAGAGAGCTGACCTAGCGGCTGACCAACGCGAGATGCCGAT 1020  
 Db 960 ATGCCCGCTCGGTGCGGAGAGAGCTGACCTAGCGGCTGACCAACGCGAGATGCCGAT 1019  
 QY 1021 GTGCTCGAGCTTGGCGGACATGCTGCGGCGGCGGCTGCGGCTGCAATTCGGCACTAGCC 1080  
 Db 1020 GTGCTCGAGCTTGGCGGACATGCTGCGGCGGCGGCTGCGGCTGCAATTCGGCACTAGCC 1079  
 QY 1081 AAAGGCTTTTGAAGCGACGAAAGGCGGCTTACTGTCGAAAGGCTGCGGCGGCGG 1140  
 Db 1080 AAAGGCTTTTGAAGCGACGAAAGGCGGCTTACTGTCGAAAGGCTGCGGCGGCGG 1139  
 QY 1141 GTGCGGCTTACGAGCGGCGGCGGAGCTGCTGAGCTGCTGCGGCGGCTGCTTACGCGGAG 1200  
 Db 1140 GTGCGGCTTACGAGCGGCGGCGGAGCTGCTGAGCTGCTGCGGCGGCTGCTTACGCGGAG 1199  
 QY 1201 CMCAGCTCG 1209  
 Db 1200 CACACGCTCG 1208

RESULT 15  
 A89751  
 LOCUS A89751 1235 bp DNA PAT 22-JAN-2000





QY 235 GGCACCCCGCCAGGTTGGGCGGAGCTGATTATTGCTCAAGTCAAGAAGCCGATA 294  
 DB 241 GGCACCCCGCCAGGTTGGGCGGAGCTGATTATTGCTCAAGGTCAGAAAGCCGATA 300  
 QY 295 GCGCGGAATACGGCCCGCTGCGACAGGGGAGATCTTGTTCAGCTTCTTGAATTTGGCC 354  
 DB 301 GCGCGGAATACGGCCCGCTGCGACAGGGGAGATCTTGTTCAGCTTCTTGAATTTGGCC 360  
 QY 355 GCGTACAGTGTTCACAGGATGCTTGTGATTCCGGCAGCAGCTAATTTGCTTACGAG 414  
 DB 361 GCGTACAGTGTTCACAGGATGCTTGTGATTCCGGCAGCAGCTAATTTGCTTACGAG 420  
 QY 415 ACCGTTCAGACCCCGGAGGCGGAGCTACCCCTGCTTCCCGGATGAGCAGAAAGTGGCCGT 474  
 DB 421 ACCGTTCAGACCCCGGAGGCGGAGCTACCCCTGCTTCCCGGATGAGCAGAAAGTGGCCGT 480  
 QY 475 CGACTCGCCCGCCAGGTTGGGCGCTTACCACTGATGCGAACCAGAGGGGCGCGGTGTG 534  
 DB 481 CGACTCGCCCGCCAGGTTGGGCGCTTACCACTGATGCGAACCAGAGGGGCGCGGTGTG 540  
 QY 535 CTGATGGGCGGCGGTGGCGGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGG 594  
 DB 541 CTGATGGGCGGCGGTGGCGGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGG 600  
 QY 595 GCGCGGCTACAGAGCGAGCGCGCATGCGCAAGCGATGGGCGGAGCGGCTTACGCTTACAG 654  
 DB 601 GCGCGGCTACAGAGCGAGCGCGCATGCGCAAGCGATGGGCGGAGCGGCTTACGCTTACAG 660  
 QY 655 ATCAACATCGCAAACTTCCGCGCACTGACAGCGGCTGCGGCGGATCCACA TCGC 714  
 DB 661 ATCAACATCGCAAACTTCCGCGCACTGACAGCGGCTGCGGCGGATCCACA TCGC 720  
 QY 715 TACATCATGCGGCTACGAGCTGAGGAGTCCCTCAAACTGCGGAGCTGCTGATTTGGGCGC 774  
 DB 721 TACATCATGCGGCTACGAGCTGAGGAGTCCCTCAAACTGCGGAGCTGCTGATTTGGGCGC 780  
 QY 775 GTCTGTGTCAGAGGCGGCGGAGCGGCAATTAATGCTTGAATTT ACTTGTGCGCTATATG 834  
 DB 781 GTCTGTGTCAGAGGCGGCGGAGCGGCAATTAATGCTTGAATTTACTTGTGCGGCAATATG 840  
 QY 835 AAACGAGTGGGCTACGAGTGTGATTTACATGACAGGCGGCGGCTTTGGAAGGCTCA 894  
 DB 841 AAACGAGTGGGCTACGAGTGTGATTTACATGACAGGCGGCGGCTTTGGAAGGCTCA 900  
 QY 895 CGACGAGCAGCTTACGAGCAGCGGAGCTTGGCGGTGACAGAGCGCTTTTACTGCGTG 954  
 DB 901 CGACGAGCAGCTTACGAGCAGCGGAGCTTGGCGGTGACAGAGCGCTTTTACTGCGTG 960  
 QY 955 GCGAATAGCCCGCTGCTGCTGCGAAGAGCTGACCTAGCGGCGTGAACAAGCGGAGATG 1014  
 DB 961 GCGAATAGCCCGCTGCTGCTGCGAAGAGCTGACCTAGCGGCGTGAACAAGCGGAGATG 1020  
 QY 1015 CCGTATGCTGAGCTTGCAGACATGGCTGCGGCGGCGGCTGCGGATTCGAAATCCGGA 1074  
 DB 1021 CCGTATGCTGAGCTTGCAGACATGGCTGCGGCGGCGGCTGCGGATTCGAAATCCGGA 1080  
 QY 1075 CTAGCCAAAGGCTTTTTCAGACGACGAGAGGGGCTTACTGCTCGAAGGGGTGGCCACGAC 1134  
 DB 1081 CTAGCCAAAGGCTTTTTCAGACGACGAGAGGGGCTTACTGCTCGAAGGGGTGGCCACGAC 1140  
 QY 1135 CTGGGGGTCGCTTACGAGAGCGGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1194  
 DB 1141 CTGGGGGTCGCTTACGAGAGCGGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200  
 QY 1195 GCGGAGCAGCTG 1209  
 DB 1201 GCGGAGCAGCTG 1215

RESULT 2  
 V49510  
 ID V49510 standard; DNA; 1260 BP.  
 AC V49510;

DT 20-OCT-1998 (first entry)  
 DE Mycobacterium sp. AlADH DNA  
 KW Alanine dehydrogenase; AlADH; ADH; diagnosis; tuberculosis; pathogen;  
 OS Swimmers disease; vaccine; epidemic; infection; identification; ss.  
 PN Mycobacterium sp.  
 PD WO9832862-A2.  
 PF 30-JUL-1998.  
 PR 29-JAN-1998: E00484.  
 PA 29-JAN-1997: EP-101339.  
 PI FLOHE L.  
 PI FLOHE L., Hutter B., Kolk A., Singh M.  
 PI WPI: 98-427958/36.  
 PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum  
 PT - used for diagnosis of tuberculosis and other mycobacterial  
 PT diseases, also for treatment and prevention, for drug screening and  
 PT for bio-transformation  
 PS Disclosure; Page 11; 57pp; German.  
 CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated  
 CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis  
 CC and other mycobacterial infections (including 'swimmers' disease', caused  
 CC by M. marinum, a fish pathogen) in humans or animals. The protein can  
 CC also be used for control of epidemics and for vaccination, to screen for  
 CC agents with anti-mycobacterial activity, and in bio-transformations that  
 CC are specific for L-alanine. Also mycobacteria can be identified by  
 CC analysis of genomic ADH sequences. ADH is an antigen that is secreted  
 CC early during infection.  
 SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 98.6%; Score 1192; DB 1; Length 1260;  
 Best Local Similarity 99.4%; Pred. No. 2, 6e-239;  
 Matches 1208; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

QY 1 ATCTGAGATTAATGAACTTCTTACACTGAAAGCTAGATGAGAGGGGTATC 60  
 DB 16 ATCTGAGATTAATGAACTTCTTACACTGAAAGCTAGATGAGAGGGGTATC 75  
 QY 61 ATGGCGCTCGGATTTCCGACGAGACCAAAACAGC-----AATTCGGGTGGCCATC 114  
 DB 76 ATGGCGCTCGGATTTCCGACGAGACCAAAACAGCAATTCCTCGGTTGGGCTATC 135  
 QY 115 ACCCGGCGCGGCGGCGGAGCAACCGGCTGCGGCGGATGAGGCTGCTATCCAGGAGGT 174  
 DB 136 ACCCGGCGCGGCGGCGGAGCAACCGGCTGCGGCGGATGAGGCTGCTATCCAGGAGGT 195  
 QY 175 GCGGAGAGGGCTGCGCTATACCGAGCGGATTTCAAGGCGGAGGCGGCAATGCTC 234  
 DB 196 GCGGAGAGGGCTGCGCTATACCGAGCGGATTTCAAGGCGGAGGCGGCAATGCTC 255  
 QY 235 GGCACGCGCGGAGGAGGTTGGGCGGAGCTGATTTAGTCAAGGTCAGAAAGCCGATA 294  
 DB 256 GGCACGCGCGGAGGAGGTTGGGCGGAGCTGATTTAGTCAAGGTCAGAAAGCCGATA 315  
 QY 295 GCGCGGAATACGGCCCGCTGCGACAGGGGAGATCTTGTTCAGCTTCTTGAATTTGGCC 354  
 DB 316 GCGCGGAATACGGCCCGCTGCGACAGGGGAGATCTTGTTCAGCTTCTTGAATTTGGCC 375  
 QY 355 GCGTACAGTGTTCACAGGATGCTTGTGATTCCGGCAGCAGTCAATTTGCTTACGAG 414  
 DB 376 GCGTACAGTGTTCACAGGATGCTTGTGATTCCGGCAGCAGTCAATTTGCTTACGAG 435  
 QY 415 ACCGTTCAGACCCCGGAGGCGGAGCTACCCCTGCTTCCCGGATGAGCAGAAAGTGGCCGT 474  
 DB 436 ACCGTTCAGACCCCGGAGGCGGAGCTACCCCTGCTTCCCGGATGAGCAGAAAGTGGCCGT 495  
 QY 475 CGACTCGCCCGCCAGGTTGGGCGCTTACCACTGATGCGAACCAGAGGGGCGCGGTGTG 534  
 DB 496 CGACTCGCCCGCCAGGTTGGGCGCTTACCACTGATGCGAACCAGAGGGGCGCGGTGTG 555  
 QY 535 CTGATGGGCGGCGGTGGCGGCGGAGCGGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 594  
 DB 556 CTGATGGGCGGCGGTGGCGGCGGAGCGGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 615







QY 895 CGACCGACCACTAGACACCCGACGCTTCGCCGACGACACGCTTTTACTGCGTG 954  
 Db 916 CGACCGACCACTAGACACCCGACGCTTCGCCGACGACACGCTTTTACTGCGTG 975  
 QY 955 GCGACATGCGCCGCTCGGTCGCGAAGACGTCGACCTACGCGCTTACCAACGCGAGATG 1014  
 Db 976 GCGACATGCGCCGCTCGGTCGCGAAGACGTCGACCTACGCGCTTACCAACGCGAGATG 1035  
 QY 1015 CCGATGCTGCTGACGCTTCGCGACGACGCTGCGCGGCGCGCTGCGCTGCAATCCGCGA 1074  
 Db 1036 CCGATGCTGCTGACGCTTCGCGACGACGCTGCGCGGCGCGCTGCGCTGCAATCCGCGA 1095  
 QY 1075 CTAGCCAAAGGCTCTTTCGACGACGAGGAGGCGTACTACTGCGAAGGAGTGGCCACGAC 1134  
 Db 1096 CTAGCCAAAGGCTCTTTCGACGACGAGGAGGCGTACTACTGCGAAGGAGTGGCCACGAC 1155  
 QY 1135 CTGGGGGTCGCTTACGACGACGCGCGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTG 1194  
 Db 1156 CTGGGGGTCGCTTACGACGACGCGCGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTG 1215  
 QY 1195 GCGGACGACGCTG 1209  
 Db 1216 GCGGACGACGCTG 1230

# RESULT 4

VA9511 standard; DNA; 682 BP.

ID V49511: 20-OCT-1998 (first entry)  
 AC V49511: 20-OCT-1998 (first entry)  
 DE Mycobacterium marinum Mar3 DNA.  
 DI Alamine dehydrogenase; AlADH; ADH; diagnosis; tuberculosis; pathogen;  
 KM swimmers disease; vaccine; epidemic; infection; identification; ss.  
 OS Mycobacterium marinum.  
 FH Key Location/Qualifiers  
 FT CDS 1..682  
 FT /tag= a  
 FT /codon\_start= 3  
 FT /product= "Mar3"  
 FT /note= "Alamine dehydrogenase"

WO9832862-A2.  
 PD 30-JUL-1998.  
 PF 29-JAN-1998; E00484.  
 PR 29-JAN-1997; EP-101339.  
 PA (FLOH/) FLOHE L.  
 PI FLOHE L., Hutter B., Kolik A., Singh M.  
 DR P-PDB: W64481.  
 PT Nucleic acid encoding alamine dehydrogenase of Mycobacterium marinum  
 PT - used for diagnosis of tuberculosis and other mycobacterial  
 PT diseases, also for treatment and prevention, for drug screening and  
 PT for bio-transformation  
 PS Claim 1: Page 34-35; 57pp; German.  
 CC This sequence encodes an alamine dehydrogenase (ADH) protein, Mar3  
 CC isolated from Mycobacterium marinum. This protein is used to diagnose  
 CC tuberculosis and other mycobacterial infections (including 'swimmers'  
 CC disease', caused by M. marinum, a fish pathogen) in humans or animals.  
 CC The protein can also be used for control of epidemics and for  
 CC vaccination, to screen for agents with anti-mycobacterial activity, and  
 CC in bio-transformations that are specific for L-alanine. Also mycobacteria  
 CC can be identified by analysis of genomic ADH sequences. ADH is an antigen  
 CC that is secreted early during infection.  
 SO Sequence 682 BP; 105 A; 254 C; 225 G; 98 T;

Query Match 38.7%; Score 467.6; DB 1: Length 682;  
 Best Local Similarity 80.4%; Pred. No. 5,4e-89;  
 Matches 548; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 98 AATTCCGGTGGGACATACCGCGCGGCTGCGGAACTAACCGCTGCGCATGAGG 157  
 Db 1 AATTCCGGTGGGACATACCGCGCGGCTGCGGAACTAACCGCTGCGCATGAGG 60

QY 158 TGCATCCAGGCGGTCGCGGAGAGGGCTCGGCTATCACGACGCGGATTTCAGGCG 217  
 Db 61 TGCATCCAGGCGGTCGCGGAGAGGGCTCGGCTATCACGACGCGGATTTCAGGCG 120  
 QY 218 CAGCGCGCAACTGTCGCGACCGCGACGAGTGTGGCGGAGCGCTGATTATGCTCA 277  
 Db 121 CCGGTCGCACTATCAGACACCGCGACGAGTGTGGCGGAGTGTGGAGCTGCTCA 180  
 QY 278 AGGTCAAGAAACCATAGCGCGGATACGCGCGCTGCGACACGCGGACATCTGTCA 337  
 Db 181 AGGTCAAGAAACCATAGCGCGGATACGCGCGCTGCGACACGCGGACATCTGTCA 240  
 QY 338 CGTCTTGTGCTGCGCGCGCTGCGACGCTGCTGCGGATGCTGTTGATTCGCGACCA 397  
 Db 241 CCTACTGCACTGCGCGCGCTGCGCGCGCTGCGACGATGCGCTGCTGATGCGGACCA 300  
 QY 398 CGTCAATTGCTAGACACGCTGCGACGCGCGGACGCGGACGCTGCTGCTGCTGCTG 457  
 Db 301 CGTCAATTGCTAGACACGCTGCGACGCGCGGACGCGGACGCTGCTGCTGCTGCTG 360  
 QY 458 TGAAGAGTGGCGGTCGCGGCTGCGCGCGGACGCGGCTGCTGCTGCTGCTGCTGCTG 517  
 Db 361 TGAAGAGTGGCGGTCGCGGCTGCGCGCGGACGCGGCTGCTGCTGCTGCTGCTGCTG 420  
 QY 518 AAGGCGCGCGGTCGCTGATGCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCG 577  
 Db 421 ACGGCGGTCGCGGCTGCTGATGCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCG 480  
 QY 578 TGATGCGCGCGGCGGACGCGCGGCTGCGACGCGGCTGCGCGGCTGCGCGGCTGCG 637  
 Db 481 TGATGCGCGGCGGACGCGCGGCTGCGACGCGGCTGCGCGGCTGCGCGGCTGCG 540  
 QY 638 CCGTACGCTTCTAGACATCAACATGCAAACTTGGCAACTGCGAGCGGATGCTGCG 697  
 Db 541 TGGTACCGCTGCTGATGATCAACATCAACAGCTGCGGACGCGGATGCTGCG 600  
 QY 698 GCGGATCCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 757  
 Db 601 GTGCGCTCGG/ CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
 QY 758 ACCTGTGATTTGGGCGCTG. 779  
 Db 661 ACATGCTGATGGGCGGCTGCT 682

# RESULT 5

N91423 standard; DNA; 1125 BP.

ID N91423: 01-FEB-1991 (first entry)  
 AC N91423: 01-FEB-1991 (first entry)  
 DE Sequence of heat-resistant alamine dehydrogenase (AH) gene with mol. wt.  
 DE 2 MD or less  
 KM Enzyme; ds.  
 OS Bacillus stearothermophilus IFO 12550.  
 PN J01043194-A.  
 PD 15-FEB-1989.  
 PF 10-AUG-1987; 200524.  
 PR 10-AUG-1987; JP-200524.  
 PA (NIRA) Unilika K.  
 DR WPI: 89-096096/13.  
 PT Recombinant plasmid for transforming Escherichia coli -  
 PT obt. by connecting heat resistant alamine dehydrogenase gene to  
 PT vector plasmid  
 PS Disclosure: Fig 2, p693; 9pp; Japanese.  
 CC A recombinant plasmid contg. heat-resistant alamine dehydrogenase (AH)  
 CC gene with mol. wt. 2 or less MD is claimed. Cells transformed with the  
 CC vector produce high levels of heat-resistant AH. Transformed E. coli  
 CC is useful as a clinical inspection reagent.  
 SO Sequence 1125 BP; 246 A; 295 C; 367 G; 217 T;

Query Match 27.9%; Score 277.2; DB 1: Length 1125;  
 Best Local Similarity 56.2%; Pred. No. 1.8e-49;

Matches 604: Conservative 0: Mismatches 458: Indels 12: Gaps 4:

```

QY 61 ATGCCGCTGCTATTCGACCGACCAACAAACAAATTCGGGTGGCCATCACCCG 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 ATGAAGATCGGATTCACAAAGAAATCAAAAACAAATGAAGAACCGGCTCCCATCATCCG 60
QY 121 GCGGGGCTGGGAAATACACCGTGTGGCATGAGTGCT---CATCGAGCAGGTGCG 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GAGAGGCTGATACGCTCTCAAGAGGGGCAATGAGTGTATGTGAGACGGAAGGC 120
QY 178 GAGAGGGCTCGCTATACACCGAGGATTTCAAGGCGCAGCGCGCACTGTGTGCG 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GCGGCTGGGTGCGGGCTTTTCGATTCGAGTATGAAGAACCGGGGCACTGTGTGCG 180
QY 238 ACCGCGACAGAGTGTGGCGCACGCTGATTTATTGCTCAAGCTCAAGAACCGATAGCG 297
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 CGAAGCTGAGAGATGCTTGAAGCGGGGAGTGTGTTAAAGTGAAGAGCGCGTGGCT 240
QY 298 GCGGATATAGCGCGGCTGCGACAGGGGAGATCTTGTGCACTTGTGATTTGGCGCG 357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 CGAGAGTTCGCTATTTTCCCGGATGATTTTGTATTTGATTTGATTTAGCCGCG 300
QY 358 TCACGTGCTTGCACGAGATCGCTTGTGATTCGCGACCGCATTCATTCCTACGAGACC 417
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GCGGAACGCGTACGAAAGCGTCTGTCGAGCAAAAAGTGTGGCATGCTTACGAGAGC 360
QY 418 GTCCAGACCGCGAGCGGCGCACTACCCCTGCTTCCCGCATGAGCAAGTCCCGGTGCA 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 GTACAGCTGCGCAAGCGGCTGCTGCGACCTGTTGACCCCATGAGTAACTCCCGGCGCC 420
QY 478 CTGCGCGCCAGTGTGGCTTACACACCTGATGCGAACCAGGCGGCGCGGTGCGTC 537
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 ATGTGCTGCAAGATGCGGCGCCAGTTTCTCGAAGAACCCCA-566GGAAGAGGCATTTTG 480
QY 538 ATGGCGGGGTGCTCCGCGCTCGAACCAGCGCTGCTGATCGCGCGCACCGCC 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 CTTCGCGGGTCCCGGAGATGGGGGCGGCAAGTACATCATCGCGCGGAGAGGGG 540
QY 598 GCGTACACGACGCGCGCATGCGCAAGGCGCATGGCGCGCACCGTTACGTTACACATC 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 GCGAGCAAGCGCGCGCAAAATCGGGTCTGCGGCGCAACGCTGACGATTTTGGACATY 600
QY 658 AACATCGAACAATTCGGAACCTGCAACGCGCGGATCTGGGCGGATCCACCTGCTAC 717
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 601 AAGCGCGACGCGTGGCGAGTCTGATGTTGTGGCGACCACTGACGACGCTCAGT 660
QY 718 TCATCGGCTTACGAGCTGAGGGTGGCGTCAAAAGTCCGACCTGTGATTTGGGCGCTC 777
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 TCCAACTGCTACCATATATCCGAGTCCGAGTCCGCGAATTCGATTTGCTCGTCCGCTC 720
QY 778 CTGTGCGCAGGCGCCAGGACCCCAATTAAGTCTGCAATTCATCTGTCCGATATGAAA 837
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 TTGATCCCGGGGCGCA---ACGGAAGTGTGAGCGAAGAGATGTTGGCTGATGACG 777
QY 838 CGAGTGGGCTACTGTTGATATAGCATGACGACAGGCGGCTGTTTCCAA---GGCTCA 894
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 778 CCGGATGCTGTGTTGCTTAATGCGCATTTGACCAAGGCGGCAATTTTCAAGACGACG 837
QY 895 CGACGACCACTTACGACCAAGCGAGTTCGCGCTGACGACAGCTGTTTACTGCTG 954
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 838 CGACGCTTACGAGAGACGATCGACATAGCTCAAGCAGCG .GTGTCATCATACGCGCTC 77
QY 955 GCGAACAATGCGCGCTGCTGAGGAGAAAGTGAAGTCTAGGCTGACCAAGCAAGAAAG 1014
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 898 GCGAACAATGCGCGC---GTGCGCGCAGCTGACATTTGCGCTTACGAACGTCAGATC 954
QY 1015 CCGATATGCTGAGCTTGCACGACATGCTGCGGCGCGCTGCGGCTCGAATCCGCGCA 1074
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 955 CCAATACGCTTGAATATGCGCAAAAGGCTACCGCGGTTGCTTGAATACCGCGGG 1014
QY 1075 CTAGCCAAAGCTTTTCAGACGACGAGAGGGGCTTACTCTCGAAGCGGTGCGC 1128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1015 CTGTTAAAGGATCAACACGCTCGACGGGCGACATGCTAGCAAGACGGTTCGCGC 1068

```

```

RESULT 6
V52155/c
ID V52155 standard; DNA; 28171 BP.
AC V52155;
DE 23-OCT-1998 (first entry)
KW Streptococcus pneumoniae genome fragment SEQ ID NO:22.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
OS computer readable medium; vaccine; pharmaceutical composition; ds.
PN MO9818931-A2.
PD 07-MAY-1998.
PF 30-OCT-1997; U19588.
PR 31-OCT-1996; US-029960.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA.
DR WPI: 98-272225/24.
PT Computer readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
PS Claim 1: Page 273-289; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded,
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridise to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 28171 BP; 8507 A; 5080 C; 5440 G; 8144 T;

Query Match 17.6%; Score 212.4; DB 1; Length 28171;
Best Local Similarity 53.9%; Pred. No. 6,5e-36;
Matches 533: Conservative 0: Mismatches 436: Indels 19: Gaps 4:
QY 61 ATGCCGCTGCTATTCGACCGACCAACAAACAAATTCGGGTGGCCATCACCCG 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21711 ATGTTAATCGGAATCCCAAAAATAAATAACGAACCGTGTGCCCTACACACT 21652
QY 121 GCGGCGTTCGCGGACACTACCGCTGTGGCATGAGGGTCTATCCAGAGTCCCGGA 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21651 GCGAGTGTTCATAGCTATGTTAGTGTGCTATCGTGTCTTATCGAAGAAATGCTG 21592
QY 181 GAGGCTGCGCTATACCGACGCGGATTTCAAGGCGCAGCGCGCAACTGCTGGCGCC 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21591 CTGGGTTCTGAGTTTACTGATGCTATCAAAAGCAAGGAGTGTGATTTGCTGAGT 21592
QY 241 GCCGACCAAGTGTGGCGGACGCTGATTTATGCTCAAGTCAAGAACCAATGACGCGCG 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21531 GCTGTGGAAGCTTTGGGCG---ACACAGTGTGTTGGAAGTAAAGATCTTTAAGTCT 21475
QY 301 GAATACGGCGGCTTCGACGAGGCGCAATCTGTTACGCTTTCGATTTGGCGCGCTCA 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21474 GAATACGGTACTTCTCGGAGAGATCTTCTCTTCCACTACTTCCATGATGCGCGTCT 21415

```

```

OY 361 CGGCTGACCGATGCGTGTGGATTCCGGCACACGCTCAATTGCCACGAGACCGTC 420
DB 21414 CCAGATTGACGATGCTATGTT-----AACGCAAAAAACAACTAACTGTT 21367
OY 421 CAGACCGCGGACGCGGCTACCTCCCTGCTGCGCCGATAGCAAGTCCCGCTGACTC 480
DB 21366 CGGACAAATCAAGCAATACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 21307
OY 481 GCGGCGGACGTTGGGCTTACCACTGATGCGCAACCAAGGGGCGGCGGCTGCTGATG 540
DB 21306 GCGTTCAATGAGAGTCTTCTTACTTAAGCAAGCTGCGCTGCTGCTGCTTACTT 21247
OY 541 GCGGCGGCTGCGGCTGCGACCGCGGAGTCTGCTGCTGCGGCGGCGGCGGCGG 600
DB 21246 GGGTGTGTACGAGGTTGCTCAAAAAGGAAAGTAACTATCATGCTGCTGCTGCT 21187
OY 601 TACACGCGACCGGCTGCGCAACGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 660
DB 21186 ACACATGCTGCGCGATCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 21127
OY 661 ATCGACAACTTCGCGACTCGACGCGGAGTCTGCGGCGGAGTCTGCTGCTGCTGCT 720
DB 21126 TCGACGCTCTCTCACTTCTAGAGAGTCTTGGAAAGTCAATTCATCTTATGCT 21067
OY 721 TCGGCTAGAGCTGAGGCTGCGCAACGCTGCGGAGGCGGCGGCGGCGGCGGCGG 780
DB 21066 AATTCATTCACATTTAGCAAGTGTGAGAGTCTGATGCTGATGCTGATGCTGCT 21007
OY 781 GTGCGAGGCGGCGGACCGCAATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
DB 21006 ATCCCTGCTGCAAAAACCGGAAATGCTGACAGATGCTGCTGCTGCTGCTGCTG 20947
OY 841 GGGGCGTACTGCTGATATAGCATGCGACGAGGCGGCGGCTTGA---AGGCTACGA 897
DB 20946 GGGTGTGTA-TGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20888
OY 898 CCAGACACTGACGACACCGGAGTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 957
DB 20887 GTACAAAGGAGATGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20828
OY 958 AACATGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1017
DB 20827 AATATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20768
OY 1018 TATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1045
DB 20767 TATATGAAAGCTTGGCTGGCAAAAGAT 20740

```

RESULT 7

V74442/C  
V74442 standard: DNA: 9280 BP.

AC V74442:  
16-MAR-1999 (first entry)  
Staphylococcus aureus contig SEQ ID #131.  
Computer readable medium: vaccine; S.aureus infection; immunodetection;  
cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
skin infection; surgical wound infection; scalded skin syndrome;  
toxic shock syndrome; ds.  
Staphylococcus aureus.  
Key Location/Qualifiers  
misc-feature 841..900  
tag-a /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

misc-feature 2641..2700  
tag-b /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

```

FT misc-feature 4441..4500
FT /tag-c
FT /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT misc-feature 6241..6300
FT /tag-d
FT /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT misc-feature 8041..8100
FT /tag-e
FT /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

```

EP-786519-42.  
30-JUL-1997.  
07-JAN-1997: 100117.  
PR 05-JAN-1996: US-009861.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA, Rosen CA;  
WPI: 97-374922/35.  
DR Polynucleotide(s) and proteins derived from Staphylococcus aureus stored on computer readable medium and used in the production of anti-S.aureus vaccines.  
PS Claim 1: Page 710-715: 3271pp: English.  
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the S.aureus DNA sequences allows putative functions to be assigned so that protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are likely to encode antigens have been identified and these polypeptides can be used in a vaccine composition against S.aureus infection. The polypeptides can also be used in a kit for the immunodetection of S.aureus in a sample. S.aureus is implicated in numerous human diseases, including cellulitis, eyelid infections, food poisoning, osteomyelitis, skin and surgical wound infections, scalded skin syndrome, toxic shock syndrome, etc. Organisms transformed with the DNA sequences can be used for recombinant production of the polypeptides. The new DNA sequences (and their fragments) are useful as primers or probes for isolating homologues of any of the S.aureus DNA sequences contained on the computer readable medium.  
CC Sequence 9280 BP: 2958 A: 1649 C: 1309 G: 3060 T:  
SO

Query Match 14.8%; Score 178.4; DB 1; Length 9280;  
Best Local Similarity 50.6%; Pred. No. 6.8e-29;  
Matches 484; Conservative 0; Mismatches 466; Indels 6; Gaps 2;

OY 151 CATGAGTGTCTATCCAGGACGATGCGGAGAGGCTGCTGCTATCCAGCGGATTC 210  
DB 9280 CATACTGTTTGTGGAACCAATTCGCGGTGATGCTTGTGAAGATGATATAC 9221  
OY 211 AAGGCGGACGCGGCGGCACTGCTGCGCACCGGACGAGTGTGGCGGACGATGTA 270  
DB 9220 AAGAGACGAGGTGCTGACATGTTGCTGACACAGCAAAAGTTGG--GATGTGATATG 9164  
OY 271 TTGCTCAAGCTCAAGAACCGATACGCGGATACGCGCTGCGACGCGGACATC 330  
DB 9163 GTATTAAAGTTAAAGAACCACTGATGATATCATATTTAAAGAGGCTGTA 9104  
OY 331 TTGTTACGCTTCTGATTTGGCGGCTGACGCTGCTTACCGATGCTTGTGATCC 390  
DB 9103 TATTTCATTAATCTTATTAAGCAATGACAAATTAATGACGCTTTATGATAGA 9044  
OY 391 GCGACCACTGATGCTTACGAGACGCTGCGACCGCGGCGGCGGCGGCGGCGGCTT 450

















GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:16:28 ; Search time 341.15 Seconds

(without alignments)  
460.653 Million cell updates/sec

Title: US-09-362-485-9

Perfect score: 1209

Sequence: 1 ATCTGCGAGTTATGCGAC.....GTTACGCCGAGCAGCAGTCG 1209

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/PTUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/1na/Backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47.6	3.9	1620	US-08-461-775-10	Sequence 10, Appl
2	47.6	3.9	28958	US-08-258-261B-6	Sequence 6, Appl
3	47.6	3.9	28958	US-08-456-837-6	Sequence 6, Appl
4	47.6	3.9	28958	US-08-457-342-6	Sequence 6, Appl
5	47.6	3.9	28958	US-08-457-646A-6	Sequence 6, Appl
6	47.6	3.9	28958	US-08-458-076A-6	Sequence 6, Appl
7	47.6	3.9	28958	US-08-764-233A-4	Sequence 4, Appl
8	47.6	3.9	28958	US-08-457-335A-6	Sequence 6, Appl
9	47.6	3.9	28958	US-08-729-214-6	Sequence 6, Appl
10	47.6	3.9	49377	US-08-764-233A-1	Sequence 1, Appl
11	47.6	3.8	2582	US-08-816-105A-2	Sequence 2, Appl
12	45.4	3.8	833	US-08-403-852D-3	Sequence 3, Appl
13	45.4	3.8	5392	US-08-403-852D-1	Sequence 1, Appl
14	42.4	3.5	1833	US-08-403-852D-6	Sequence 6, Appl
15	42.4	3.5	2185	US-08-173-508-3	Sequence 3, Appl
16	42.4	3.5	2185	US-08-265-310-3	Sequence 3, Appl
17	41.2	3.4	2414	US-08-461-775-11	Sequence 11, Appl
18	41.2	3.4	2668	US-07-642-734C-3	Sequence 7, Appl
19	41.2	3.4	20235	US-08-804-227C-7	Sequence 7, Appl
20	40.8	3.4	44377	US-08-804-198-1	Sequence 1, Appl
21	40.8	3.4	44377	US-08-804-198-1	Sequence 1, Appl
22	40.6	3.4	4257	US-08-690-473-1	Sequence 1, Appl
23	40.6	3.4	12001	US-08-458-568A-11	Sequence 11, Appl
24	40.2	3.3	2064	US-08-343-428-1	Sequence 1, Appl
25	40.2	3.3	11219	US-07-642-734C-1	Sequence 1, Appl
26	39.6	3.3	474	US-08-403-852D-14	Sequence 14, Appl
27	39.6	3.3	1524	US-08-461-775-10	Patent No. 5512669-1

28	39.6	3.3	1525	7	5229279-1	Patent No. 5229279
29	39.4	3.3	459	4	US-08-387-942C-35	Sequence 35, Appl
30	39.4	3.3	1998	7	5212296-8	Patent No. 5212296
31	39.4	3.3	12588	4	US-08-387-942C-1	Sequence 1, Appl
32	39.2	3.2	420	1	US-08-470-179-148	Sequence 148, App
33	39.2	3.2	2353	6	PCT-US92-06840-1	Sequence 1, Appl
34	39.2	3.2	43280	3	US-08-804-227C-1	Sequence 1, Appl
35	38.8	3.2	8051	4	US-08-576-626A-2	Sequence 2, Appl
36	38.8	3.2	11219	2	US-07-642-734C-1	Sequence 1, Appl
37	38.6	3.2	1187	2	US-08-440-856A-2	Sequence 2, Appl
38	38.6	3.2	3231	2	US-08-074-121-4	Sequence 4, Appl
39	38.6	3.2	3231	6	PCT-US94-06447-4	Sequence 4, Appl
40	38.4	3.2	1215	4	US-08-947-726A-1	Sequence 1, Appl
41	38.4	3.2	2048	5	US-08-776-251-1	Sequence 1, Appl
42	38.2	3.2	2109	4	US-08-555-568B-20	Sequence 20, Appl
43	38	3.1	30001	1	US-08-125-468-1	Sequence 1, Appl
44	38	3.1	30001	1	US-08-474-933-1	Sequence 1, Appl
45	37.8	3.1	2588	3	US-08-796-414B-6	Sequence 6, Appl

## ALIGNMENTS

RESULT 1  
US-08-461-775-10

; Sequence 10, Application US/08461775

; Patent No. 5858773

; GENERAL INFORMATION:

; APPLICANT: MAZODIER, Philippe

; TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESS: Burns, Doane, Swecker & Mathis

; STREET: George Mason Bldg., Washington & Prince Sts.

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/461.775

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/050,313

; FILING DATE: 10-MAY-1993

; APPLICATION NUMBER: FR 9011186

; FILING DATE: 10-SEP-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Crane-Feuery, Sharon E

; REGISTRATION NUMBER: 36,113

; REFERENCE/DOCKET NUMBER: 010830-035

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-6620

; TELEFAX: (703) 836-2021

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1620 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..1620

; US-08-461-775-10

Query Match 3.9%; Score 47.6; DB 3; Length 1620;  
 Best Local Similarity 45.7%; Pred. No. 0.0046;  
 Matches 321; Conservative 0; Mismatches 369; Indels 12; Gaps 4;

QY 114 CACCCCGGCGGCGTGGGAACTAACCCGCTGAGTGGTCCATCCAGCAG 173  
 DB 879 CACCTCACCAGGCTGACCTGATGCCAGAGAGTGGGCTTCAAGCTGACAGGCGCG 938  
 QY 174 TCGCGAGAGAGGCTGCGCTATACCGAGCGGATTTCAAGCGCGCGCAATGCT 233  
 DB 939 TCGAGAGTGTGGGACCGCGCGCTCAACCGTCAAGAGAGAGAGAGAGAGAGAT 998  
 QY 234 CCGGACCGCGGACAGAGTGTGGGCGGAGCTATTTATGCTCAAGTCAAGAGAGAGAT 293  
 DB 999 GAGAGCGGCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1058  
 QY 294 AGCGGCGGAGATAGCGGCGGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 353  
 DB 1059 CAGGCTGAT 1118  
 QY 354 C-GCGTACAGTGTGCTGACCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 412  
 DB 1119 CCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1178  
 QY 413 AGACGCTCAGAT 470  
 DB 1179 GAGAGCTGAT 1238  
 QY 471 CCGTCACTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 527  
 DB 1239 CCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1298  
 QY 528 CCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 585  
 DB 1299 CCGCGAT 1358  
 QY 585 -CGCGCGAT 641  
 DB 1359 GATCGCGAT 1418  
 QY 642 TAGGCTTGTGAGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 701  
 DB 1419 CAAGGCGGAT 1478  
 QY 702 GATCGAT 761  
 DB 1479 CAGCGAT 1538  
 QY 762 GGTGATTTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 803  
 DB 1539 GCTCTGAT 1580

RESULT 2  
 US-08-258-261B-6

Sequence 6, Application US/08258261B  
 Patent No. 5639949  
 GENERAL INFORMATION:  
 APPLICANT: Schupp, Thomas  
 APPLICANT: Ligon, James M.  
 APPLICANT: Beck, James Joseph  
 APPLICANT: Hill, Dwight Steven  
 APPLICANT: Ryals, John Andrew  
 APPLICANT: Gatiney, Thomas Deane  
 APPLICANT: Lam, Stephen Ting  
 APPLICANT: Hammer, Phillip E.  
 APPLICANT: Umes, Scott Joseph  
 TITLE OF INVENTION: Genes for the synthesis of  
 TITLE OF INVENTION: antipathogenic substances  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ciba-Geigy Corporation

STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10532  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: IBM PC compatible  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/258,261B  
 FILING DATE: 08-JUN-1994  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/457,205  
 FILING DATE: 01-JUN-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Elmer, James Scott  
 REGISTRATION NUMBER: 36,129  
 REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
 TELEPHONE: 919-541-8614  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 28958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 US-08-258-261B-6

Query Match 3.9%; Score 47.6; DB 1; Length 28958;  
 Best Local Similarity 44.1%; Pred. No. 0.01;  
 Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

QY 148 GCGCATGAGTGTCTATCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 207  
 DB 25306 GCGCAGAT 25365  
 QY 208 TTCAAGCGGAT 267  
 DB 25366 TTGAT 25425  
 QY 268 TTATGTCTCAAGTGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 327  
 DB 25426 TCGCTGCGGAT 25485  
 QY 328 ATCTTGTACGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 387  
 DB 25486 AAGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 25545  
 QY 388 TCGGAT 447  
 DB 25546 CTTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 25605  
 QY 448 CTTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 505  
 DB 25605 CAGCTGAT 25665  
 QY 506 TATGCGAAGAT 565  
 DB 25666 CCGCTGAT 25725  
 QY 566 CCGAGG-TGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 624  
 DB 25726 CGCTGCTGAT 25785  
 QY 625 GCGATGAT 684

Accession	Sequence	Position
Db	25786 GCGTTGCGTGCAGGGCGCTCTCTCCCTCGAGAGCGGGCCCCCATCGCCCTCTGGCGAC	25845
Oy	685 GCCGAGTTCTGGGGCGGATCCACTCGCTACTCATCGGCGCTACGAGGTGAGSGTGC	744
Db	25846 AAGGCGCTACACCGTCGCGCGGCAACGGGGGCGCATGGCGCGCGTGGAGTGGGCGCTCC	25905
Oy	745 GTCAACAGTCCGACCTGGTGTATGGGGCGTCTGGTCCACAGGGCG	792
Db	25906 GACCTCCAGACTACTCTGCTCTCCCTGGGGGCGACAGGCTTCATCGCC	25953

## RESULT

```

US-08-456-837-6
: Sequence 6, Application #US/08456837
: Patent No. 5643774
:
: GENERAL INFORMATION:
:
: APPLICANT: Schnupp, Thomas
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
:
: ZIP: 10532
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.0
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/456,837
: FILING DATE: 01-JUN-1995
:
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-JUN-1994
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
:
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
:
: US-08-456-837-6

```

Query Match	3.98;	Score 47.6;	DB 1;	Length 20958;
Best Local Similarity	44.18;	Pred. No. 0.01;		
Matches 286;	Conservative	0;	Mismatches 359;	Indels 3;
				Gaps 2

[illegible]

## RESULT

```

05-08-457-342-6
1 Sequence 6, Application US/08457342
2 Patent No. 5662898
3
4 GENERAL INFORMATION:
5
6 APPLICANT: Schupp, Thomas
7
8 APPLICANT: Ligon, James M.
9
10 APPLICANT: Beck, James Joseph
11
12 APPLICANT: Hill, Dwight Steven
13
14 APPLICANT: Ryals, John Andrew
15
16 APPLICANT: Gathney, Thomas Deane
17
18 APPLICANT: Lam, Stephen Ting
19
20 APPLICANT: Hammer, Phillip E.
21
22 APPLICANT: Uknes, Scott Joseph
23
24 TITLE OF INVENTION: Genes for the synthesis of
25
26 TITLE OF INVENTION: antipathogenic substances
27
28 NUMBER OF SEQUENCES: 22
29
30 CORRESPONDENCE ADDRESS:
31
32 ADDRESS#1: Ciba-Geigy Corporation
33
34 STREET: 7 Skyline Drive
35
36 CITY: Hawthorne
37
38 STATE: NY
39
40 COUNTRY: USA
41
42 ZIP: 10532
43
44 COMPUTER READABLE FORM:
45
46 MEDIUM TYPE: Floppy disk
47
48 COMPUTER: IBM PC compatible
49
50 OPERATING SYSTEM: PC-DOS/MS-DOS

```



OY	268	TTTTGGCTCAAGGTAAAGAAACCGATACGGGGGGAATACGGCGGCTCTCGACACGGGAG	327
Db	25426	TGCGCTCGCCCAAGACAAACCCGCCCGAGACACGCTCTCTGGACGGAGGAAAGCAAGGAC	25485
OY	328	ATCTTGTTCAGCTTCTTGCAATTTGGCCGGCTACGTGCTTGCACCGATGCGTTTGGAT	387
Db	25486	AAGCTCGTCTTCTGCTTCTTCTTCGGGGAAGGCTTCGAGTGGGAAGGATGGCCCTTCGGTG	25545
OY	388	TCCGGCACACGATCAATTGAGCTACGAGACGCTCCAGACCGCCGAGGGGGGCACTACCCCTG	447
Db	25546	CTCGACTCTCGCCCGCTCTTCGCGCTCAGCTCGAAGCATGGGAGCGGCGGTGCGCTCT	25605
OY	448	CTTGCCCCGATGAGCGAAGTGGCCGGGTGCACTGCGCCCGCAGGTTGGGCTTACCAC--C	505
Db	25606	CACGTCGAGTGAAGCTCTGCTGCGCTCTGCGCCCGACGAGAGGGGCCCCCTCCCTGCAC	25665
OY	506	TGATCCGACCAACCAAGGGGGCGCGGTGATGATGGGGGGGTGGCCGAGCTCGAAGACGG	565
Db	25666	GCGCTCGACGTCGTACAGCCCGCCCTCTTTGCCGCTAATGATCTCCCTGGCGGCCCTCTGG	25725
OY	566	CCGAGC-TCGTGATGATGGGCGCGGACCGCGGCTACAAACGACGCCGATCGCCCAAC	624
Db	25726	GCGTCGTCGGGCTGAGAGCCCGCCCGCTGCTGGCCACATCGAGGGGAGATTCGCGGC	25785
OY	625	GGCATGGGGCGCACCGGTATACGGTTGTAGCATCAACATGACAAATTCGGGCACTGCAC	684
Db	25786	GCTTCGTGGCAGGGCTCTTCTCCCTCGAGGACGGGCGCCGCAATGCGCGCCCTGGCCACG	25845
OY	685	GCCGAGTCTCGGGCGGATCCACACTCGCTACTATCGCCCTACGAGCTCGAGGGGTGC	744
Db	25846	AAAGGCTCACCACGTCGCGCGCAACGAGGGGCATGAGCGCGCGTAGAGTCTCGCGCCCTCC	25905
OY	745	GTCAAACGCGCGACCTGTTAATTTGGGGCGCTCGTGGTGCACAGGGGCC 792	
Db	25906	GACCTTCAGACTACCTGCTCTCCCTGGGGCGAGAGCTTTCATGCCC 23593	

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

? FILING DATE: 01-JUN-1995  
 ? APPLICATION NUMBER: 08/258,261  
 ? FILING DATE: 08-Jun-1994  
 ? ATTORNEY/AGENT INFORMATION:  
 ? NAME: Elmer, James Scott  
 ? REGISTRATION NUMBER: 36,129  
 ? REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
 ? TELECOMMUNICATION INFORMATION:  
 ? TELEPHONE: 919-541-8614  
 ? TELEFAX: 919-541-8689  
 ? INFORMATION FOR SEQ ID NO: 6:  
 ? SEQUENCE CHARACTERISTICS:  
 ? LENGTH: 28958 base pairs  
 ? TYPE: nucleic acid  
 ? STRANDEDNESS: single  
 ? TOPOLOGY: linear  
 ? MOLECULE TYPE: DNA (genomic)  
 ? HYPOTHEetical: NO  
 ? ANTI-SENSE: NO  
 ?

Query Match	3.9%	Score 47.6	DB 2	Length 28958
Best Local Similarity	44.1%	Pred. No. 0	0.0	
Matches 286	Conservative	0	Mismatches 359	Indels 3
			Gaps	2
QY 148	GCCATGATGATGCTTCATCCAGCAGCAGTGTCCCGAGAGGGCTCGGCTATATCAGCAGCGGAT	207		
Db 25306	GCCACGACGACACTACCTCGCGGATGTGGCGTATTGCGTGGCCACACCGCGGCCAC	25365		
QY 208	TTCAAGGCGGACGAGCGCCCAACTGTGTGGGACCGCCGACAGTGTGGGCGGACGCTAT	267		
Db 25366	TTGCGACACCGGCGCGCTCTCGTACCCACAAACCGCGACAGCTCTCTCGCGCTGCAC	25425		
QY 268	TTATTGTCAGGTCAAAGAAACCATACGCGGGGGAATACGGCGCGCTCGACAGCGGAG	327		
Db 25426	TCGCTGCGCCAGACGACCCCGCCGACACCGCTCTCGGACGGAGGGAACCCAGCGC	25485		
QY 328	ATCTTGTTACAGTTCCTTCATTTGGCCCGTCACGTGCTTGCACCGATGCGTTGTGAT	387		
Db 25486	AAGCTCGTCTGCTTCTTCTCTGGCAAGGCTGCGATGGAAAGGATAGCCCTCTCGGTG	25545		
QY 388	TCGCGCACCGTCATTTCCCTACGAGACCGTCCAGACCGCGGACGCGGCTACCCCTG	447		
Db 25546	CTGCACTCCTCCCGCTCTTCGCGCTCAGCTCGAAGCATGCGACGCGCGCTCGCTCT	25605		
QY 448	CTTGCCCCGATGAGGCANATGCTGTGTGCTGCTGCGCGCCGCGGAGTTGGGCTTACAC	-C 505		
Db 25606	CACGTCGATGAGTGCCTCTCCGCTCTCTGCGCGCGAGAGGCGCCCTCCTCGAC	25665		
QY 506	TGATCGAACCACAAAGGGGCGCGGTGTGCTGATGGCGGGGTGCGCGGCTGCAACCG	565		
Db 25666	CGCGTCGAGTGTACAGCCGCCCTCTTGGCGTCAATGTTCTCCTTGGGCGCCTCTGG	25725		
QY 566	CCGACG-TCGTGTATCGCGCGCGGACCGCGGCTACACGACGCGCGCATCGCCAC	624		
Db 25726	CGCTGCTGGGGTGAAGCGCGCGCGCGCTGTCGCGCACATGACGAGGAGATCGCGCC	25785		
QY 625	GGCATGGGCGGACCGTTACGCTTCTAGACATCAACATGACCAAACTTGGCACTCGAC	684		
Db 25786	GGCTTGTGTCGAGGGGCTCTCTCCCTCGAGAGCGGGCGCGCGATCGCGCCTGGCGAC	25845		
QY 688	GCAGATGTGAGGCGGATTCACACACTCGCTACTATGAGGCTGACAGGCTGACAGGATG	744		
Db 25846	AAAGGCTCACACGATCGCGGCAACGGGGGCGCATAGCGCGCGTGAAGTGTGGCGCTCC	25905		
QY 745	GTCAAAAGTGGCGGACCTGTGATTGGGGCGGTCTGTGTCCACAGGCGC 792			
Db 25906	GACCTTCAGACCTACTCTCCTCTGGGGCGACAGGCTCTCCATGCGC 25953			

RESULT 7

US-08-764-233A-4





```

? TELEPHONE: 919-541-8614
? TELEFAX: 919-541-8689
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 28938 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEetical: NO
? ANTI-SENSE: NO
US-08-457-335A-6

```

Query Match	3.9%	Score 47.6;	DB 2;	Length 28958;
Best Local Similarity	44.1%	Pred. No. 0.01;		
Matches 286;	Conservative	0;	Mismatches 359;	Indels 3;
				Gaps 2

QY	148	GGCATTGAGAGTCTCATCCAGGCAGGTGCGGAGAGAGGGGTGCGCTATACCCAGCGGGAT	207
Db	25306	GCCACGAGACCTACCTCGCGGATGTGGCTTATTCGTGGCCACACCCGCGCCAC	25365
QY	208	TTCAAGGGGCGTAGCGCGCAGTGGTTCGGCACCGCCGACTAGGTGGACCCACCTGAT	267
Db	25366	TTTCAGGCAACCGGCGGCTCTCTGTACCCCAACCCGCAAGACCTCTCTCCGCGCTGCAC	25425
QY	268	TTATTGCTCAGGTCAAGAACCCGATTAGCGCGGAATAGCGCCCTCGGCACAGGGCAG	327
Db	25426	TGCGTCGCGCCAGACAGAGCCGCCCGGAGACACGTCCTCGAGCAGGAGGAAACCAAGCC	25485
QY	328	ATCTGTTTCAGCTTCTGGATTTTGGCCGGGTCACGAGTGTGCACGATGCGATTGTTGAT	387
Db	25486	AAGCTGCTTCTTGTCTTTTCTGGGCAAGGCTTCGACGTGGAGAAGGATGGCCCTCTGCTG	25545
QY	388	TCCGGCACCACGCTCATTTGCTTCACAGAGACCGTCCAGACCGCCGACGCGCCTACCCCTG	447
Db	25546	CTCGACTCTGCGCCGCTCTTCGGCGCTCAGTCTGAAAGCATGCGAGCGCGGCTCGGCTCT	25605
QY	448	CTTGCCCCGATAGCGAACTCGCCGCTCGACTTCGCGCCAGTTGGCGCTTACCAAC--C	505
Db	25606	CACGTCGATGAGGCTGCTTCGCCGCTTCGCGCCCGAGAGAGGGGCCCTCCCTCGAC	25665
QY	506	TGATGCGAACCCAAAGGGGCGCGGCTGTCTGATGGCGGGGTGCGCGGCGTCGAACCG	565
Db	25666	CGCGTCGAGGTGTACAGCCCGCCTCTTTGCCGTCATGTCTCCCTGGCGGCTCTGTGG	25725
QY	566	CCGAGCGTCGTGTGATCGGCGCGGACCGCGGCTACACGACGACCCCGATCCGCCAAC	624
Db	25726	CGCTGCGTGGGCTGTAGAGCCCGCGCGCTGTCTGGGCAACGTACAAGGAGATCGCGCC	25785
QY	625	GGCATGGGCGGCAACGTTACGCTTTCAGACATTCACATATGACAAACATTGGCACTCGAC	684
Db	25786	GGCTTCGTGCGAGGGGCTTCTCCCTCGAGAGACGGGCGCCGATCGCGCCCTCGCGCAC	25845
QY	685	GCCGATTTGCGGCGCGATTCACATCCACTCTACTATCGGCTACGACCTCGAGGGTGGC	744
Db	25846	AAAGGGCTACACACGTCGCGGCAACAGGGGCGATAGGCGCGCGCTGTGAGGCTCGGCGCTC	25905
QY	745	GTCAAACGTGCGGACCTGATGATTGGGGCGGCTCTGTGTCCACAGGGGCC792	
Db	25906	GACCTTCAGACTTACTGCTCCTCTGGGGGAGACGGCTTCCATCGCC25953	

RESULT 9  
US-08-729-214-6  
; Sequence 6, Application US/08729214  
; Patent No. 5817502  
; GENERAL INFORMATION:  
; APPLICANT: Ligon, James M.  
; APPLICANT: Hill, Dwight Steven  
; APPLICANT: Ryals, John Andrew  
; APPLICANT: Hammer, Phillip E.  
; APPLICANT: van Pee, Karl-Helz

1 APPLICANT: Kimer, Sebline  
 2 TITLE OF INVENTION: Genes for the synthesis of  
 3 antipathogenic substances  
 4 TITLE OF INVENTION: antipathogenic substances  
 5 NUMBER OF SEQUENCES: 27  
 6 CORRESPONDENCE ADDRESS:  
 7 ADDRESSEE: Ciba-Geigy Corporation  
 8 STREET: 520 White Plains Road  
 9 City: Tarrytown  
 10 STATE: NY  
 11 COUNTRY: USA  
 12 ZIP: 10591  
 13 COMPUTER READABLE FORM:  
 14 MEDIUM TYPE: Floppy disk  
 15 COMPUTER: IBM PC compatible  
 16 OPERATING SYSTEM: PC-DOS/MS-DOS  
 17 SOFTWARE: Patent in Release #1.0, Version #1.25  
 18 CURRENT APPLICATION DATA:  
 19 APPLICATION NUMBER: US/08/729,214  
 20 FILING DATE: TBA  
 21 CLASSIFICATION: 435  
 22 ATTORNEY/AGENT INFORMATION:  
 23 NAME: Melgs, J. Timothy  
 24 REGISTRATION NUMBER: 38,241  
 25 REFERENCE/DOCKET NUMBER: CGC 1506/CIPS  
 26 TELECOMMUNICATION INFORMATION:  
 27 TELEPHONE: 919-541-8587  
 28 TELEFAX: 919-541-8689  
 29 INFORMATION FOR SEQ ID NO: 6:  
 30 SEQUENCE CHARACTERISTICS:  
 31 LENGTH: 28958 base pairs  
 32 TYPE: nucleic acid  
 33 STRANDEDNESS: single  
 34 TOPOLOGY: linear  
 35 MOLECULE TYPE: DNA (genomic)  
 36 HYPOTHEetical: NO  
 37 ANTI-SENSE: NO  
 38  
 39 US-08-729-214-6

Query Match	3.98;	Score 47.6;	DB 2;	Length 28958;
Best Local Similarity	44.18;	Pred. No. 0.01;		
Matches 286;	Conservative	0;	Mismatches 359;	Indels 3;
				Gaps 2;

OY	148	GGCATTGAGTGGTTCATTATCAGGACGAGCGCCGAGAGGGCTCGCGTATATCAACGACGGGAT	207
Db	25306	GCCACGACGACCTTACACCTCTCCGATATGCGCTATTCCGTGGCCACACCCCGCGCCAC	25365
OY	208	TTCAAGCGCGCGCGGCCACATGCGTCGACCGCCGACGACAGTGTGGGCCGACGCTGAT	267
Db	25366	TTCCGAGCACCGGCGCGCTCTCTGTAAGCCACAAACCGCGAGACCTCTTCTCCGCGTCCGAC	25425
OY	268	TTATTGCTCAAGGTCAAGAAAGAACCGATTAGCGGGGGAATACGGCGCGCTCGACACGGGGG	327
Db	25426	TGCGTCGCGCCAGACAGACGCCGCCCGGAGACCGTCTCTCGACAGGAGCGGAAGCCACGGC	25485
OY	328	ATCTTGTTACGCTCTTGCAATTGGCCGCGTCAGCTGCTTGACCCAGTCGTTGGAT	387
Db	25486	AAGCTCGCTTCTGTTCTTCTCTGGCAAGCTCGCATGTGGAGGGATGACCTCTCGCTG	25445
OY	388	TCCGGCACCAAGTCAATTGGCCATGACGACACCGTCGACACCGCCGAGGGGCACTACCCTG	447
Db	25546	CTCGACTCTTCGCCGCTCTTCGCCGCTCAGCTCGAAGCATGCGACCGGCGCTCGCTCT	25605
OY	448	CTTGCCCCGATGAGCAAGTCGCGCGGTGACTCGCGCCGACGTTGGGCTTACAC--C	505
Db	25606	CACGTCGAGTGAAGCTCTCGCCGCTCTCGGCCGCGAGAGAGGGGCCCTCTCTCTGAC	25665
OY	506	TGATCGAACCCAAAGGGGCGCGGCTGTGCTATGAGCGGGGTGCCCGGCTCGAACCGG	565
Db	25666	CGCGTCGACGTCGTACAGCGCCGCTCTTTGCGGTATGATGTTCCCTGGCGGCCCTCTGG	25725
OY	566	CCGACG--TGCTGTGATGCGCGCCGCGCACCGCGGCTTACACGACGCCGCGATCGCCAAC	624

```

Db 25726 CGCTGCTGCGGCTAGAGCCGCGCGCGGCGCAAGTCAGAGGCGAGATCGCGCG 25785
OY 625 GGCATGGCGCGACCGCTTACGTTTACATCAATCAATCGCAACTCGAC 684
Db 25786 GCGTTGCTGCGAGGCGCTCTCTCCGAGGAGCGGCGCGCGCGCTGCGCAGC 25845
OY 685 GCGGAGTTGCGGCGCGATCCACTGCTCTATCGCTTACGCTCAGAGTCCAGGCTGCG 744
Db 25846 AAGCGCTACCAACCGTCCCGCGCAAGCGGCGCGCGCTGCGCGCTGCGCGCTCC 25905
OY 745 GTCAACGTCGACCTGCTGATTTGGGCGCTCTGCTGCTGCGCGCGCG 792
Db 25906 GACCTCCAGACCTACCTGCTCTCCGCGCGAGAGGCTCTCCATCGCC 25953

RESULT 10
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Snezana
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Weigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 1506/CIP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Sorangium cellulosum
; IMMEDIATE SOURCE:
; CLONE: p98/1, pL3, and pVKM15
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 383..760
; OTHER INFORMATION: /product- "Sorb"
; OTHER INFORMATION: /note- "This gene encodes a protein that is highly homologous

```

```

; OTHER INFORMATION: the reductase domains of type I PKs such as eryA from
; OTHER INFORMATION: Saccharopolyspora erythraea."
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 927..19874
; OTHER INFORMATION: /product- "Sora"
; OTHER INFORMATION: /note- "Gene product is highly homologous to type I PKs."
; OTHER INFORMATION: are known to be involved in the synthesis of polyketide
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 942..7115
; OTHER INFORMATION: /product- "Module 1 of Sora"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 7203..12884
; OTHER INFORMATION: /product- "Module 2 of Sora"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 13455..19616
; OTHER INFORMATION: /product- "Module 3 of Sora"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 19871..46318
; OTHER INFORMATION: /product- "Sorb"
; OTHER INFORMATION: /note- "Gene product is highly homologous to type I PKs ge
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 19870..24556
; OTHER INFORMATION: /product- "Module 1 of Sorb"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 24638..30820
; OTHER INFORMATION: /product- "Module 2 of Sorb"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 30881..3544
; OTHER INFORMATION: /product- "Module 3 of Sorb"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 35528..40114
; OTHER INFORMATION: /product- "Module 4 of Sorb"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 40190..46318
; OTHER INFORMATION: /product- "Module 5 of Sorb"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 46851..47891
; OTHER INFORMATION: /product- "Sorb"
; OTHER INFORMATION: /note- "The protein encoded by the sorb gene is highly
; OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
; OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the
; OTHER INFORMATION: polyketide rapamycin."
; US-08-764-233A-1

Query Match 3.9%; Score 47.6; DB 2; Length 49377;
Best Local Similarity 44.1%; Pred. NO. 0.012; Mismatches 339; Indels 3; Gaps 2;
Matches 286; Conservative 0;

OY 148 GGCATGAGTGTCTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 207
Db 41609 GCCACAGACGACCTCACCCTCGGATGTGGCTTGTGCTGCGCCACACCGGCGCAC 41668
OY 208 TTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTAT 267
Db 41669 TTCGACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGAC 41728
OY 268 TTATTGCTCAAGGTCAAGAACCGATAGGCGGAGGAATAGGCGGCGGCGGCGGCGGCGGAG 327
Db 41729 TCGCTGCGGCGGAGACAGCGCGCGCGCGGAGACCGGCTCTCTGAGGAGGCGGAGGAGCGAGC 41788

```

QY 328 AACTGTCACGCTTCTGATTTGGCCGGTCACGCTTGCACCGCATGCTGTGAT 387  
 DB 41789 AAGCTGCTCTGCTCTCTCTGCGCAAGCTCCAGTGGAGAGATGCGCTCTGCG 41848  
 QY 388 TCCGACACACGTCATTTGCTACGAGACCGTCCAGACCGCGCGCATACCTG 447  
 DB 41849 CTCGATCCCTCGCGCTCTCTGCGCTACGCTGAGACATGACGCGCGCTCTCT 41908  
 QY 448 CTGCCCCCATGAGCGAAGTGGCGGCTGACCTGCGCGCGCGCTTGCCTTCCAC -C 505  
 DB 41909 CAGCTGAGAGAGCTGCTGCGCTCTGCGCGCGCGCGCGCGCTCTCTCTGAC 41968  
 QY 506 TGATGCAACCCAAAGGGGGCGCGGTGTGCTGATGGCGGGGTGCGCGCAACCG 565  
 DB 41969 CGGCTGAGAGCTGACAGCGCGCTTGTGCGGTATGCTCTCTGCGCGCTCTG 42028  
 QY 566 CCGAGC-TGCTGCTGATGCGCGCGCGCGCGCGCTACACGACCGCGCTCTCAAC 624  
 DB 42029 CGCTGCTGCGCTGAGAGCGCGCGCGCTGCTGCGCGCGCGCGCGCTCTCT 42088  
 QY 625 GGGATGGGCGCGACCGCTTACGCTTACATGACATGACAACTGCGCAACTGAC 684  
 DB 42089 GCGTCTGCTGAGAGCGCTCTCTCTGAGAGAGCGCGCGCGCTGCGCGAGC 42148  
 QY 685 GCGAGTCTGCGCGCGCGATGACACTGCTACTCATGCGCTTACGAGCTGAGGCTG 744  
 DB 42149 AAGCGCTGACACCGCTGCGCGCGCGCGCGCGCGCGCTGCGCGCTCTCT 42208  
 QY 745 GTCAAGTCCGACCGCTGATGGGGCGCTGCTGCTGCGCGCGCGCTCTCTCT 792  
 DB 42209 GACCTCGACCTACTCTGCTCTCTCTGCGCGCGCGCTCTCTCTCTCTCT 42256

## RESULT 11

US-08-816-105A-2  
 : Sequence 2, Application US/08816105A  
 : Patent No. 598982  
 : GENERAL INFORMATION:  
 : APPLICANT: Criegee, Anneliese; Dellweg, Hans-Georg; Lenz,  
 : APPLICANT: Jurgen; Schroder, Werner; Pape, Hermann;  
 : APPLICANT: Goetz, Klaus; Schaper, Beate; Hemker, Michael;  
 : APPLICANT: Piepersberg, Wolfgang; Distler, Jurgen;  
 : APPLICANT: Straumann, Ansgar  
 : TITLE OF INVENTION: PROCESSES FOR PREPARING ACARINOSTYL  
 : TITLE OF INVENTION: TRANSFERASE AND FOR USING IT IN THE  
 : TITLE OF INVENTION: CONVERSION OF ACARBOSE HOMOLOGUES  
 : TITLE OF INVENTION: INFO ACARBOSE, FOR THE PREPARATION  
 : TITLE OF INVENTION: OF ACARBOSE HOMOLOGUES  
 : NUMBER OF SEQUENCES: 16  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Sprung Kramer Schaefer & Briscoe  
 : STREET: 660 White Plains Road  
 : CITY: Tarrytown  
 : STATE: New York  
 : COUNTRY: USA  
 : ZIP: 10591-5144  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage  
 : COMPUTER: Apple Macintosh 6500  
 : OPERATING SYSTEM: System 7.5  
 : SOFTWARE: Mordperfect 3.5  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/816,105A  
 : FILING DATE: 14-MAR-1997  
 : CLASSIFICATION: 435  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: DE 19625269.5  
 : FILING DATE: 25-JUN-1996  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: DE 19611252.4  
 : FILING DATE: 22-MAR-1996  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Kurt G. Briscoe

REGISTRATION NUMBER: 33,141  
 REFERENCE/DOCKET NUMBER: Bayer 9814-KGB  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (914) 332-1700  
 TELEFAX: (914) 332-1844  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2582 nucleotides  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 US-08-816-105A-2

Query Match 3.8% Score 46; DB 4: Length 2582;  
 Best Local Similarity 47.3%; Pred. No. 0.013;  
 Matches 139; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 353 CCGCTGACGCTCTTGCACCGATGCGTGTGATTCGGGACACGCTCATGCTTACG 412  
 DB 1576 CCGCGCTGCGCGCGCGCGCGCGCTTACGAGCTGCTGCGCGCGCGCTTACG 1635  
 QY 413 AGACGCTGACACCGCGCGCGCGCGCGCGCTTACGAGCTGCTGCGCGCGCGCT 472  
 DB 1636 TGACGCTGAGAGAGCGCGCGCGCGCGCGCGCTTACGAGCTGCTGCGCGCG 1695  
 QY 473 GTGACTGCGCGCGCGCGCGCGCTTACGAGCTGCTGCGCGCGCGCGCTTACG 532  
 DB 1696 GCGAGTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1755  
 QY 533 TGCTATGGCGGGGGGTCGCCCGCTGACACCGCGCGCGCTGCTGCTGCTGCTG 592  
 DB 1756 TCGGCGGGGCTGCTGACCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCT 1815  
 QY 593 CGCGCGCTGACAAAGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTG 646  
 DB 1816 CCGCGGAAACGCTGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCT 1869

## RESULT 12

US-08-403-852D-3  
 : Sequence 3, Application US/08403852D  
 : Patent No. 5891695  
 : GENERAL INFORMATION:  
 : APPLICANT: Blanc, Veronique  
 : APPLICANT: Blanche, Francis  
 : APPLICANT: Crouzet, Joel  
 : APPLICANT: Jacques, Nathalie  
 : APPLICANT: Lacroix, Patricia  
 : APPLICANT: Thibaut, Denis  
 : APPLICANT: Zagorec, Monique  
 : APPLICANT: Debussche, Laurent  
 : APPLICANT: De Crey-Lagard, Valerie  
 : TITLE OF INVENTION: Polypeptides Involved in The  
 : TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
 : TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
 : NUMBER OF SEQUENCES: 43  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
 : STREET: 1300 I Street, N.W., Suite 700  
 : CITY: Washington  
 : STATE: D.C.  
 : COUNTRY: USA  
 : ZIP: 20005-3315  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patentin Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/403,852D  
 : FILING DATE: 10-MAY-1995  
 : PRIOR APPLICATION DATA:



APPLICANT: Crouzet, Joel  
APPLICANT: Jacques, Nathalie  
APPLICANT: Lacroix, Patricia  
APPLICANT: Thibaut, Denis  
APPLICANT: Zagorec, Monique  
APPLICANT: Debussche, Laurent  
APPLICANT: De Crecy-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved In The  
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finegan, Henderson, Farbow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,852D  
FILING DATE: 10-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806, 0054-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1833 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.pristinaespiralis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 103..1689  
US-08-403-852D-6

Query Match 3.5%; Score 42.4; DB 3; Length 1833;  
Best Local Similarity 50.5%; Pred. No. 0.087;  
Matches 103; Conservative 0; Mismatches 01; Indels 0; Gaps 0;

533 TCTGATGGCGGGGTCGCCGCGTGAACCGCGACGTCGTGATGCGGCGCGCA 592  
DB 413 TGTATGCTCGAGCTGGGCTCTTGGCGCGCGGCTCCGCTGGGGGCGCTTCGCCCAAGACT 472

593 CGCGCGCTACCAAGAGCGCCGCGATGCCAAGCGCATGGCGCGCGACGTTAGGCTTAG 652  
DB 473 CGCGCCAACTATCGCGCGCGCGCGCGATGGCGCTGGCGCGCGCGCTGCTGGCGACCA 532

653 ACATCAACATCGACCAACTTGGCACTGACCGCGCGAGTTCGCGCGCGGATCCAACTAC 712  
DB 533 CCACCTCCCGCATCATGACAGCTTTCGACGACGACGACGACGACGACGACGACGACG 592

713 GCTACTCATCGCGCTACGAGCTCG 736

DB 593 TGTGGGAGCGCGCCGCTACTGCG 616

RESULT 15  
US-08-173-508-3  
Sequence 3, Application US/08173508  
Patent No. 5616485  
GENERAL INFORMATION:  
APPLICANT: Bartfield, Daniel  
APPLICANT: Butler, Michael J.  
APPLICANT: Hadary, Dany  
APPLICANT: Jenish, David  
APPLICANT: Krieger, Timothy  
TITLE OF INVENTION: STREPTOMYCIN PROTEINASES AND IMPROVED  
TITLE OF INVENTION: STREPTOMYCIN STRAINS FOR EXPRESSION OF PEPTIDES AND  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/173,508  
FILING DATE: 23-DEC-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 18740/125/CACO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202 672 5300  
TELEFAX: 202 672 5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2185 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 531..2069  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 531..902  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 903..2069  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 531..533  
OTHER INFORMATION: /note="Met at position -124  
US-08-173-508-3

Query Match 3.5%; Score 42.4; DB 1; Length 2185;  
Best Local Similarity 47.4%; Pred. No. 0.091;  
Matches 127; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

525 CGCGGTGCTGATGGCGGGGTCGCCAAGCGCGCGCGCTGCTGATGATGCG 584  
DB 1250 CGCGCTCGCATGACCGCGGTGCGACCGCGGTGAAGGTGGCGCGCATCAAGTCTCCAA 1309

585 CGCGGACCGCGCGCTACACGACCGCGCATCGCCAAAGCGCATGGCGGACCGCTTAC 644

```

Db 1310 CCCCACGGCTTCTCTACACCGAGCCGCTGCTCGCGCTTCACTGCGCGCCGAGCA 1369
QY 645 GGTCTAGACATCAACATCGAACCTTCGGCACTGACGCCGAGTTCTGCGCGCGAT 704
Db 1370 CGCGCTCGACGTGACCAACAGCTATTACACGACCCGCTGCTACTTCAACTGCAAGGA 1429
QY 705 CCACACTGCTACTCATGCGCTAGAGCTGAGAGGTGCGCTCAACGTCGCCGACTGT 764
Db 1430 CGACCCCGACGAGAGCGCTGCTGAGAGCCGCTCTCGCGGCGCTCCGCTAGCGGAGAA 1489
QY 765 GATTGGGCGCTCTGCTGCTGCCAGCGCC 792
Db 1490 GAAGGCGCGGTCAACGTGCGCGCGCC 1517

```

Search completed: June 22, 2000, 15:18:59  
 Job time: 18617 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:07:56 ; Search time: 5541.94 Seconds  
(without alignments)  
884.230 Million cell updates/sec

Title: US-09-362-485-9  
Perfect score: 1209  
Sequence: 1 ATCTGCAGATTAATCGAAC.....GTTACGCCGAGCNCACGTGG 1209

Scoring table: IDENTITY\_NDC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 202611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:  
1: em\_est1.\*  
2: em\_est2.\*  
3: em\_est3.\*  
4: em\_est4.\*  
5: em\_est5.\*  
6: em\_est6.\*  
7: em\_est7.\*  
8: em\_est8.\*  
9: em\_est9.\*  
10: em\_est10.\*  
11: em\_est11.\*  
12: em\_est12.\*  
13: em\_est13.\*  
14: em\_est14.\*  
15: em\_est15.\*  
16: em\_est16.\*  
17: em\_est17.\*  
18: em\_est18.\*  
19: em\_est19.\*  
20: gb\_est1.\*  
21: gb\_est2.\*  
22: gb\_est3.\*  
23: gb\_est4.\*  
24: gb\_est5.\*  
25: gb\_est6.\*  
26: gb\_est7.\*  
27: gb\_est8.\*  
28: gb\_est9.\*  
29: gb\_est10.\*  
30: gb\_est11.\*  
31: gb\_est12.\*  
32: gb\_est13.\*  
33: gb\_est14.\*  
34: gb\_est15.\*  
35: gb\_est16.\*  
36: gb\_est17.\*  
37: gb\_est18.\*  
38: gb\_est19.\*  
39: gb\_est20.\*  
40: gb\_est21.\*  
41: gb\_est22.\*  
42: gb\_est23.\*  
43: gb\_est24.\*  
44: gb\_est25.\*

45: gb\_est26.\*  
46: gb\_est27.\*  
47: gb\_est28.\*  
48: gb\_est29.\*  
49: gb\_est30.\*  
50: gb\_est31.\*  
51: gb\_est32.\*  
52: gb\_est33.\*  
53: gb\_est34.\*  
54: gb\_est35.\*  
55: gb\_est36.\*  
56: gb\_est37.\*  
57: gb\_est38.\*  
58: gb\_est39.\*  
59: gb\_est40.\*  
60: gb\_est41.\*  
61: gb\_est42.\*  
62: gb\_est43.\*  
63: gb\_est44.\*  
64: gb\_est45.\*  
65: gb\_est46.\*  
66: gb\_est47.\*  
67: gb\_est48.\*  
68: gb\_est49.\*  
69: gb\_est50.\*  
70: gb\_est51.\*  
71: gb\_est52.\*  
72: gb\_est53.\*  
73: gb\_est54.\*  
74: gb\_est55.\*  
75: gb\_est56.\*  
76: gb\_est57.\*  
77: gb\_est58.\*  
78: gb\_est59.\*  
79: gb\_est60.\*  
80: gb\_est61.\*  
81: gb\_est62.\*  
82: gb\_est63.\*  
83: gb\_est64.\*  
84: gb\_est65.\*  
85: gb\_est66.\*  
86: gb\_est67.\*  
87: gb\_est68.\*  
88: gb\_est69.\*  
89: gb\_est70.\*  
90: gb\_est71.\*  
91: gb\_est72.\*  
92: gb\_est73.\*  
93: gb\_est74.\*  
94: gb\_est75.\*  
95: gb\_est76.\*  
96: gb\_est77.\*  
97: gb\_est78.\*  
98: gb\_est79.\*  
99: gb\_est80.\*  
100: gb\_est81.\*  
101: gb\_est82.\*  
102: gb\_est83.\*  
103: gb\_est84.\*  
104: gb\_est85.\*  
105: gb\_est86.\*  
106: gb\_est87.\*  
107: gb\_est88.\*  
108: gb\_est89.\*  
109: gb\_est90.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result

Query

No.	Score	Match Length	DB	ID	Description
C 1	57.4	4.7	925	82	CNS0091P
C 2	55.2	4.6	935	82	CNS0091P
C 3	51.8	4.3	955	74	AV390505
C 4	50.2	4.2	844	82	CNS0052P
C 5	49	4.1	910	82	CNS0060N
C 6	47	3.9	846	82	CNS010RT
C 7	46.6	3.9	1101	83	CNS0181E
C 8	46.4	3.8	935	82	CNS006XK
C 9	46.2	3.8	645	82	CNS01213
C 10	45.8	3.8	935	82	CNS006XK
C 11	45.4	3.8	591	69	AW128786
C 12	45.2	3.7	744	83	CNS017ZK
C 13	45.2	3.7	1101	83	CNS017SY
C 14	45	3.7	315	51	AW171257
C 15	45	3.7	846	82	CNS010RT
C 16	44.8	3.7	932	82	CNS0072Q
C 17	44	3.6	843	82	CNS00CS1
C 18	43.8	3.6	427	51	AW136678
C 19	43.8	3.6	512	74	AV396794
C 20	43.8	3.6	521	74	AV397116
C 21	43.8	3.6	526	74	AV394827
C 22	43.6	3.6	1101	83	CNS016E3
C 23	43.2	3.6	439	81	QNS084063
C 24	43.2	3.6	798	82	CNS008A7A
C 25	43	3.6	413	74	AV369647
C 26	41.8	3.5	414	80	AW286708
C 27	41.8	3.5	465	80	AW287007
C 28	41.8	3.5	530	74	AV396353
C 29	41.8	3.5	839	82	CNS004AB
C 30	41.6	3.5	932	82	CNS0072Q
C 31	41.6	3.4	506	74	AW180713
C 32	41.6	3.4	870	82	CNS006E2
C 33	41.4	3.4	1101	82	CNS000LX
C 34	41.4	3.4	1201	83	CNS014B7
C 35	41.4	3.4	639	60	AW182194
C 36	41.2	3.4	774	105	AQ327321
C 37	41.2	3.4	910	82	CNS0060N
C 38	41	3.4	605	46	AW461529
C 39	41	3.4	1101	83	CNS0175Y
C 40	40.8	3.4	384	64	AW054773
C 41	40.8	3.4	385	21	D48746
C 42	40.6	3.4	437	74	AV395790
C 43	40.6	3.4	506	74	AV396405
C 44	40.4	3.3	574	91	AW048186
C 45	40.2	3.3	916	83	CNS0155Q
C 1					AL053013 Drosophill
C 2					AL053013 Drosophill
C 3					AV390505 AV390505
C 4					AL056652 Drosophill
C 5					AL0565629 Drosophill
C 6					AL099337 Drosophill
C 7					AL108764 Drosophill
C 8					AL056051 Drosophill
C 9					AL101589 Drosophill
C 10					AL056051 Drosophill
C 11					AW128786 fea7705.y
C 12					AL108698 Drosophill
C 13					AL1108460 Drosophill
C 14					AW171257 605069D09
C 15					AL099337 Drosophill
C 16					AL066742 Drosophill
C 17					AL059666 Drosophill
C 18					AW136678 sb-2b04.y
C 19					AV396794 AV396794
C 20					AV397116 AV397116
C 21					AV394827 AV394827
C 22					AL107216 Drosophill
C 23					AW0864063 nbep0022M
C 24					AL055081 Drosophill
C 25					AV396947 AV396947
C 26					AW286804 LG1_261-B
C 27					AW287007 LG1_264-C
C 28					AV396353 AV396353
C 29					AL054280 Drosophill
C 30					AL056742 Drosophill
C 31					AW180713 MGA0863f
C 32					AL064221 Drosophill
C 33					AL078845 Drosophill
C 34					AL103945 Drosophill
C 35					AW182194 605086F10

## ALIGNMENTS

	<p>collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org/The%20BDGP/Drosophila_melanogaster/BAC_library.html">http://www.fruitfly.org/The BDGP Drosophila melanogaster</a>. A BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the Escherichia coli strain yz<sup>-</sup> cm bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://hpac.med.buffalo.edu/drosophila_bac.htm">http://hpac.med.buffalo.edu/drosophila_bac.htm</a>.</p>					
FEATURES						
source	1..925 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPCT-98" /clone="BACRID16" /note="end : TET3"					
BASE COUNT	120 a        61 c        61 g        172 t        511 others					
ORIGIN						
	Query Match                       4.7%; Score 57.4; DB 82; Length 925; Best Local Similarity 11.9%; Pred. No. 0.058; Matches 43; Conservative 175; Mismatches 142; Indels 0; Gaps 0;					
Dy	442	CCCCGTCCTCCCGATGAGCGAAGTGGCGGGTGACTGCCGCCCAAGTGTGGCGCTTAC	501			
		::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::				
Db	920	SCGCSGSBSBSSSSSMWTSSSNSRBCSSSBSSSTSSMSBSSBSBSSSSSGTS	861			
Oy	502	CACCTGATGCACAACCAGAAGGGGCCCGCGGTGTGATGCGGGGGGTCCCGCGCTGA	561			
		::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::				
Db	860	SACVKNASSSCGCGGMABCMCSSSSSCGSASABRYKTRAGAKRRGGSGASA	801			
Oy	562	CCGGCCGACGTGCTGTGATCGCGCGCGCACCGCCGCGCTACACGACGCCATGCC	621			
		::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::				
Db	800	SHSSSSACBSSSSSCSASCMSASSSSASRSRSGGAGGAGSSSRSSSSSSASGS	741			
Oy	622	AACGGCATGGCGCGCACCCTTAGCGTTCTGTGACATCATGACAAACTTCGGCACTC	681			
		::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::				
Db	740	VYSSASSSSSSCSSSVSCSYAASMCSBSSSSASBASSSSSSSASACASCCCTSMSC	681			
Oy	682	GACGCCGAGTCTCGCGCGCGATCCACACTCGCTACTCATCGGCTACGAGCTCGAGGT	741			
		::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::				
Db	680	SCSTASMSAARSSSSSSSCSSMSAMSSASASSSSSSSGSSSSGSAACGSMSSG	621			

RESULT	1
CNS0091P/c	
LOCUS	
DEFINITION	CNS0091P 925 bp DNA GSS 03-JUN-1999
ACCESSION	Drosophila melanogaster genome survey sequence TERT3 end of BAC #
VERSION	BACR18D16 of RPI-98 library from Drosophila melanogaster (fruit
KEYWORDS	fly), genomic survey sequence.
SOURCE	AL053013
ORGANISM	AL053013.1 GI:4934461
	GSS.
	fruit fly.
	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 925)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
	Bp 191 91006 EVRR cedex - FRANCE (E-mail : sequefegenoscope.cns.fr
COMMENT	- Web : www.genoscope.cns.fr determination of this BAC-end sequence was carried out as part of a

RESULT	2
CNS0091P	
LOCUS	CNS0091P 925 bp DNA
DEFINITION	CNS0091P Drosophila melanogaster genome survey sequence TE13 end of BAC #
	BACR19D16 of RPCT-98 library from Drosophila melanogaster (fruit
	fly), genomic survey sequence.
ACCESSION	AL053013
VERSION	AL053013.1 GI:4934461
KEYWORDS	GSS:
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 925)



Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source  
Location/Qualifiers  
1. 844  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACR1421"  
/note="end : TET3"

BASE COUNT 261 a 112 c 92 g 35 t 344 others  
ORIGIN

Query Match 4.2%; Score 50.2; DB 82; Length 844;  
Best Local Similarity 17.2%; Pred. No. 1.4;  
Matches 65; Conservative 161; Mismatches 150; Indels 1; Gaps 1;

257 CCAGCGCTGATTATGCTCAAGTCACAAAGACGATACCGCGGAAATACGCCCTGCG 316  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Db 768 CCMVCCCTTTTHTTTCMYAMMRYAMMMAMMMVMMVMMVMMVMMVMMVMMVMMVMM 709  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Qy 317 GACACGGGAGATCTGTCACGCTCTGCTGATTTGGCCGCTACGCTGCTGACCGATG 376  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Db 708 NNYCBHSCBHTTTCNTCTVSSBSBHTCTHNSNTTNTTNTTNTTNTTNTTNTTNT 649  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Qy 377 CGTTGTGATTCGGGACGACGCTCAATTCCTCAGAGACCCTCCAGACCGCCGAGCGGG 436  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Db 648 CYCCGCCCTTCGCBYBNTSCSYYSBSCGSSSSSSSSSSSSSSSSSSSSSSSSSS 589  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Qy 437 CACTAC-CCCTGCTGCCCCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 495  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Db 588 CGCTSBVCSBHTTTCNTCTVSSBSBHTCTHNSNTTNTTNTTNTTNTTNTTNTTNT 529  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Qy 496 GCTTACACGCTGATGACGACGACGACGACGACGACGACGACGACGACGACGACG 555  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Db 528 SSCBYCBSCBHTTTCNTCTVSSBSBHTCTHNSNTTNTTNTTNTTNTTNTTNTTNT 469  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Qy 556 GTGACACGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 615  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Db 468 CBSBSCGSSSSSSSTKCCSSGSSSTKCCSSGSSSTKCCSSGSSSTKCCSSGSS 409  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Qy 616 ATCGGCAACGCGATGG 632  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Db 408 KRTSTSTSSSSGCGTG 392  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :

RESULT 5  
CNS0060N/c 910 bp DNA GSS 03-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #  
DEFINITION BACR1421 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION AL065629.1 GI:4944698  
VERSION AL065629.1  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 910)  
GENOSCOPE.  
AUTHORS Direct Submission  
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : [secref@genoscope.cns.fr](mailto:secref@genoscope.cns.fr))  
COMMENT - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org>. The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Oseguwa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source  
Location/Qualifiers  
1. 910  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACR1421"  
/note="end : T7"

BASE COUNT 202 a 63 c 112 g 198 t 335 others  
ORIGIN

Query Match 4.1%; Score 49; DB 82; Length 910;  
Best Local Similarity 20.5%; Pred. No. 2.4;  
Matches 47; Conservative 97; Mismatches 85; Indels 0; Gaps 0;

Qy 393 CACGACGCTAATGCTCAAGTCACAAAGACGATACCGCGGAAATACGCCCTGCTGCG 452  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Db 713 SRVMSAGSS 654  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Qy 453 CCGGATGAGCGAGACGCGGCTGCTGATGAGGCGGCGGCTGCTGACGCTGATGCG 512  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Db 653 VAAVAAVSSSSVAVSS 594  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Qy 513 AACCCAGGCGGCGGCTGCTGATGAGGCGGCGGCTGCTGACGCGGCTGACGCGG 572  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Db 593 CASSSSSSGCGGSSVCCSS 534  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Qy 573 CGTGTGATGCGGCGGCGGCTGCTGATGAGGCGGCGGCTGCTGACGCGGCTGACG 621  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :  
Db 533 CRSCGCGCCGSS 485  
11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :

RESULT 6  
CNS010R/c 845 bp DNA GSS 26-JUL-1999  
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC  
DEFINITION BACN04N13 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION AL099337.1 GI:5610948  
VERSION AL099337.1  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 846)  
GENOSCOPE.  
AUTHORS Direct Submission  
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : [secref@genoscope.cns.fr](mailto:secref@genoscope.cns.fr))  
COMMENT - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
<http://www.edgp.ebi.ac.uk>. This Drosophila melanogaster BAC  
library (DrosBAC) was made by Alain Billand at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector











## ORIGIN

Query Match 3.7%; Score 45; DB 51; Length 315;  
Best Local Similarity 50.7%; Pred. No. 12;  
Matches 108; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 391 GGCACACGCTCAATGCTCAGAGACCGTCCAGACCGCGGAGGGGCGACTACCGCTGCTT 450  
DB 58 GGCATGGCGCTCTCCATCCAAAGACAACAGCTGGTGCGCGACCAACTTTAAACACACCC 117  
QY 451 GCGCCGATGAGCGAAGTGGCCGCTGACCTGCGCCGCGGCGCTTACCCACTGATG 510  
DB 118 GCGCTAAGAGACGCAAGGGGCGGCTCAGCGCGATAGCGGCGGCGGCTCGCAGCAGCAG 177  
QY 511 CGAACCCAGAGGGGCGCGGCTGCTGATGGCGGGGTCGCCGCGTGCAGCCGCGCAG 570  
DB 178 CGTGGGCGACGGGCGCGCGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 237  
QY 571 GTCGTGTGATGGCGCGCGCGCGCGCTAC 603  
DB 238 GTGGCGCGT 36CGGCGCGCGCGCGCTCTCC 270

## RESULT 15

CNS010RJ 846 bp DNA GSS 26-JUL-1999  
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC  
DEFINITION BACN04N13 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION AL099337  
VERSION AL099337.1 GI:5610948

## KEYWORDS

## SOURCE

## ORGANISM

Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 846)

## REFERENCE

AUTHORS  
TITLE  
JOURNAL

## COMMENT

Direct Submission  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.

## FEATURES

## SOURCE

Location/Qualifiers  
1..846

/organism="Drosophila melanogaster"

/plasmid="pBelobAC11"

/db\_xref="taxon:7227"

/clone\_lib="DrosBAC"

/clone="BACN04N13"

/note="end : 77"

BASE COUNT 300 a 47 c 40 g 175 t 284 others

## ORIGIN

Query Match 3.7%; Score 45; DB 82; Length 846;  
Best Local Similarity 16.2%; Pred. No. 14;  
Matches 38; Conservative 109; Mismatches 88; Indels 0; Gaps 0;

QY 506 TGATCGGAACCCAGGGGCGCGGCTGCTGATGGCGGGGTCGCCGCTGCAACCG 565  
DB 574 KGAGSGRRARSSAAGAGGGGCGGCGGSSSSSSSSSSSSSSSSSSSSSSSSSSSS 633  
QY 566 CCGACGTGTGTGATCGCGCGCGCGCGCGCTACAAAGCCGCGCATGCCCAAG 625

DB 634 SSGSSSSAAASAGASSSVSSASAVSSSSSSSASASASASVSSSSSVSAVAASSSS 693  
QY 626 GCATGGCGCGCGCGCTTACGGTTCTAGACATCAACATCGACAACTTGGCACTGACG 685  
DB 694 SSVSVSSSSAAVAAVAVASASACASVSSSVSGCAASASASTTSSSSSASSSSSG 753  
QY 686 CCGAGTTCTGCGCGCGGATCCACACTCGCTACTCATGCGCCTACGAGCTCGAGGG 740  
DB 754 SSASGASSSSSAAASAAASVSAVAVASASSSSSSASSSSSSSSSSSSS 808

Search completed: June 22, 2000, 12:08:02  
Job time: 10396 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:57:02 ; Search time 8627.09 Seconds

(without alignments)  
-139.371 Million cell updates/sec

Title: US-09-362-485-10

Perfect score: 1236

Sequence: 1 ATCTTGACAGATTATCGAAC.....GGAGAGCAGATGTCGGCC 1236

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 862769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : GenBml:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl1:\*  
27: em\_pl2:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_hc91:\*  
33: gb\_hc92:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_hc93:\*  
42: gb\_hc94:\*  
43: gb\_hc95:\*  
44: gb\_hc96:\*

45: gb\_hc97:\*  
46: em\_hc91:\*  
47: em\_hc92:\*  
48: em\_hc93:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pr5:\*  
52: gb\_hc98:\*  
53: gb\_hc99:\*  
54: gb\_hc10:\*  
55: gb\_hc911:\*  
56: gb\_hc912:\*  
57: gb\_hc913:\*  
58: gb\_hc914:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1236	100.0	1236	5	A87613	A87613 Sequence 10
2	1236	100.0	1236	5	A89753	A89753 Sequence 10
3	1234.4	99.9	1237	5	A87607	A87607 Sequence 4
4	1234.4	99.9	1237	5	A89747	A89747 Sequence 4
5	1234.4	99.9	56414	1	MTV002	AL008967 Mycobacte
6	1232	99.7	1235	5	A87506	A87606 Sequence 3
7	1232	99.7	1235	5	A89746	A89746 Sequence 3
8	1226.4	99.2	1228	5	A87608	A87608 Sequence 5
9	1226.4	99.2	1228	5	A89748	A89748 Sequence 5
10	1223	98.9	1235	5	A87609	A87609 Sequence 6
11	1223	98.9	1235	5	A89749	A89749 Sequence 6
12	1222.4	98.9	1235	5	A87611	A87611 Sequence 8
13	1222.4	98.9	1235	5	A89751	A89751 Sequence 8
14	1220	98.7	1245	5	A87605	A87605 Sequence 2
15	1220	98.7	1245	5	A89745	A89745 Sequence 2
16	1220	98.7	1260	5	A87604	A87604 Sequence 1
17	1220	98.7	1260	5	A89744	A89744 Sequence 1
18	1220	98.7	1260	5	A89744	A89744 Sequence 1
19	1215.4	98.3	1229	5	A87610	X63069 M.tuberculo
20	1215.4	98.3	1229	5	A89750	A87610 Sequence 7
21	1208	97.7	1209	5	A87612	A89750 Sequence 7
22	1201	97.2	1208	5	A89752	A87612 Sequence 9
23	1194	96.6	1194	2	MT092472	A89752 Sequence 9
24	698	56.5	720	5	A89759	U92472 Mycobacteri
25	538.4	43.6	40745	1	SC151	A89759 Sequence 16
26	467.6	37.8	682	5	A89758	AT109848 Streptomy
27	467.6	37.8	682	5	A89758	A89758 Sequence 15
28	377	30.5	8046	2	AF049107	A89752 Sequence 19
29	351.2	28.4	11514	2	AE002028	AF049107 Myxococcu
30	331.4	26.8	2780	1	AE002028	AE002028 Delnococt
31	330.8	26.8	1585	2	AE070716	AJ23818 Rhizobium
32	312.8	25.3	1736	1	BACALDH	AF070716 Vibrio pr
33	293	23.7	1477	2	AF070715	M33299 B.stearothe
34	288.4	23.3	2113	2	D37807	AF070715 Shewanell
35	277.2	22.4	1125	5	E01847	D37807 Phormidium
36	258	20.9	137740	1	D90900	E01847 DNA encodin
37	243.2	19.7	1375	1	AB013821	D90900 Synecocyst
38	243	19.7	1776	1	BACALDH	AB013821 Enterobac
39	235.6	19.1	1709	1	BACALDH	M33298 B.sphaericu
40	235.6	19.1	12316	1	BS282015	L20916 Bacillus su
41	235.6	19.1	217420	1	BS282015	282015 B.subtilis
42	218.2	17.7	1714	2	AF070714	299120 Bacillus su
43	107.2	8.7	10512	2	AE001565	AF070714 Carnobact
44	99.2	8.0	10512	2	AE000639	AE001565 Helicobac
45	74.2	6.0	3160	1	RR05294	AE000639 Helicobac
						U05294 Rhodospiril

## ALIGNMENTS





```

QY 601 TACAAAGCAGCCCGCATGCGCAAGGCGATGGCGCGCGATTACGTTCTAGACATCAAC 660
DB 601 TACAAAGCAGCCCGCATGCGCAAGGCGATGGCGCGCGATTACGTTCTAGACATCAAC 660
QY 661 ATGCAAACTTGGCACTCGACGCGGAGTTCTGCGCGCGATCCACTGCTACTCA 720
DB 661 ATGCAAACTTGGCACTCGACGCGGAGTTCTGCGCGCGATCCACTGCTACTCA 720
QY 721 TCGGCGCTAGAGCTGAGGCTGCGGCAAGGCGGATGCTGATGTTGGGCGCGCTG 780
DB 721 TCGGCGCTAGAGCTGAGGCTGCGGCAAGGCGGATGCTGATGTTGGGCGCGCTG 780
QY 781 GTGCGAGCGCGCAAGGCGCAAAATTAATGCTGCAATTCATTGTCGCGCATAGAAACA 840
DB 781 GTGCGAGCGCGCAAGGCGCAAAATTAATGCTGCAATTCATTGTCGCGCATAGAAACA 840
QY 841 GGTGGCGTACTGCTGATATAGCATCGACAGGCGGCTGTTTTCGAAGCTCAGACCG 900
DB 841 GGTGGCGTACTGCTGATATAGCATCGACAGGCGGCTGTTTTCGAAGCTCAGACCG 900
QY 901 ACCACCTAGACAGCAGCGAGTTCGCGCGTACGACAGCAGCTGTTTACTGCGTGGCAAC 960
DB 901 ACCACCTAGACAGCAGCGAGTTCGCGCGTACGACAGCAGCTGTTTACTGCGTGGCAAC 960
QY 961 ATGCGCGCTCGGTGCGGAGAGAGAGCTGACCTAGCGCTGACCAAGCGAGATGCCGTAT 1020
DB 961 ATGCGCGCTCGGTGCGGAGAGAGAGCTGACCTAGCGCTGACCAAGCGAGATGCCGTAT 1020
QY 1021 GTGCTGACACTTGGCGAGATGCTGCGGCGGCGGTGCGGATGCGGACTAGGC 1080
DB 1021 GTGCTGACACTTGGCGAGATGCTGCGGCGGCGGTGCGGATGCGGACTAGGC 1080
QY 1081 AAGGCTTTGCAAGCAGCAGAGGCGGTACTGTCCGAAGGCGGTGCGCAAGCAGCTGGG 1140
DB 1081 AAGGCTTTGCAAGCAGCAGAGGCGGTACTGTCCGAAGGCGGTGCGCAAGCAGCTGGG 1140
QY 1141 GTGCGCTTACGAGAGCGCGCGACGCTGCTGCTGACTCTGCGCGCTGTTACGCGGAG 1200
DB 1141 GTGCGCTTACGAGAGCGCGCGACGCTGCTGCTGACTCTGCGCGCTGTTACGCGGAG 1200
QY 1201 CACAGCTCGGAGTAAAGGAGGAGATGATGCGGC 1236
DB 1201 CACAGCTCGGAGTAAAGGAGGAGATGATGCGGC 1236

RESULT 4
LOCUS A89747 1237 bp DNA
DEFINITION Sequence 4 from Patent WO9832862.
ACCESSION A89747
VERSION A89747.1 GI:6738281
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1237)
AUTHORS
FLOHE, L. and Singh, M.
TITLE
L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL
Patent: WU 9832862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAJAY (DE)
FEATURES
source
1..1237
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 394 c 386 g 221 t
ORIGIN
Query Match 99.9%; Score 1234.4; DB 5; Length 1237;
Best Local Similarity 99.9%; Pred. No. 1.6e-169;
Matches 1235; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATCTTGCAATTATGAACTTTCTTACACTGAGCGCTACGATTCAGAGGGGTAAATC 60

```

```

DB 1 ATCTTGCAATTATGAACTTTCTTACACTGAAAGCTACAGATGAGAGGGGTATC 60
QY 61 ATGCGCGTGGTATTCGACCGAGACCAAAACAAACGAATTCGGGTGGCCATACCCCG 120
DB 61 ATGCGCGTGGTATTCGACCGAGACCAAAACAAACGAATTCGGGTGGCCATACCCCG 120
QY 121 GCCGCGTGGGGAATTAACCGTGTGGCCATGAGGTGCTCATCAGGAGGTGCCGA 180
DB 121 GCCGCGTGGGGAATTAACCGTGTGGCCATGAGGTGCTCATCAGGAGGTGCCGA 180
QY 181 GAGGCGTGGCTATCAGCAGCGCGGATTTCAAGGCGGAGCGCGCAACTGTGGCACC 240
DB 181 GAGGCGTGGCTATCAGCAGCGCGGATTTCAAGGCGGAGCGCGCAACTGTGGCACC 240
QY 241 GCCGACAGGTGTGGGCGCGAGCTGATTTATGCTCAAGGTCAAAAGACCATAGCGCG 300
DB 241 GCCGACAGGTGTGGGCGCGAGCTGATTTATGCTCAAGGTCAAAAGACCATAGCGCG 300
QY 301 GAATAGCGCGCGCTGCGACAGCGGAGATCTTGTTCAGCTTCTGATTTGGCGCGTCA 360
DB 301 GAATAGCGCGCGCTGCGACAGCGGAGATCTTGTTCAGCTTCTGATTTGGCGCGTCA 360
QY 361 GGTGCTTGACCGATGCGTGTGATTCGCGACACGTCATTTGCTATAGACCGTTC 420
DB 361 GGTGCTTGACCGATGCGTGTGATTCGCGACACGTCATTTGCTATAGACCGTTC 420
QY 421 CAGACCGCGAGCGGCGCACTACCCCTGCTTGGCCCGATGAGCAGGAATGGCGGTGACTC 480
DB 421 CAGACCGCGAGCGGCGCACTACCCCTGCTTGGCCCGATGAGCAGGAATGGCGGTGACTC 480
QY 481 GCCGCGCAGGTGTGGCTTACCTGATGCGAACCCAAAGGCGGCGGCTGCTGATG 540
DB 481 GCCGCGCAGGTGTGGCTTACCTGATGCGAACCCAAAGGCGGCGGCTGCTGATG 540
QY 541 GCGGCGGTGCGCGGCTGGAACCGGCGCAGCTGCTGATGCGCGCGCGCGCGGCG 600
DB 541 GCGGCGGTGCGCGGCTGGAACCGGCGCAGCTGCTGATGCGCGCGCGCGCGGCG 600
QY 601 TACAGCGAGCGCGGATGCGCAAGGCGATGGGCGCGGAGCTTACGCTTACAGATCAAC 660
DB 601 TACAGCGAGCGCGGATGCGCAAGGCGATGGGCGCGGAGCTTACGCTTACAGATCAAC 660
QY 661 ATGCAAACTTGGCACTCGACGCGGAGTCTGCGCGCGATCCACTGCTACTCA 720
DB 661 ATGCAAACTTGGCACTCGACGCGGAGTCTGCGCGCGATCCACTGCTACTCA 720
QY 721 TCGGCGCTAGAGCTGAGGCTGCGGCAAGGCTGCTCAAGCTGCTGATGTTGGGCGCTCTG 780
DB 721 TCGGCGCTAGAGCTGAGGCTGCGGCAAGGCTGCTCAAGCTGCTGATGTTGGGCGCTCTG 780
QY 781 GTGCGAGCGCGCAAGGCGCAAAATTAATGCTGCAATTCATTGTCGCGCATAGAAACA 840
DB 781 GTGCGAGCGCGCAAGGCGCAAAATTAATGCTGCAATTCATTGTCGCGCATAGAAACA 840
QY 841 GGTGGCGTACTGCTGATATAGCATCGACAGGCGGCTGTTTTCGAAGCTCAGACCG 900
DB 841 GGTGGCGTACTGCTGATATAGCATCGACAGGCGGCTGTTTTCGAAGCTCAGACCG 900
QY 901 ACCACCTAGACAGCAGCGAGTTCGCGCGTACGACAGCAGCTGTTTACTGCGTGGCAAC 960
DB 901 ACCACCTAGACAGCAGCGAGTTCGCGCGTACGACAGCAGCTGTTTACTGCGTGGCAAC 960
QY 961 ATGCGCGCTCGGTGCGGAGAGAGAGCTGACCTAGCGCTGACCAAGCGAGATGCCGTAT 1020
DB 961 ATGCGCGCTCGGTGCGGAGAGAGAGCTGACCTAGCGCTGACCAAGCGAGATGCCGTAT 1020
QY 1021 GTGCTGACACTTGGCGAGATGCTGCGGCGGCGGTGCGGATGCGGACTAGGC 1080
DB 1021 GTGCTGACACTTGGCGAGATGCTGCGGCGGCGGTGCGGATGCGGACTAGGC 1080
QY 1081 AAGGCTTTGCAAGCAGCAGAGGCGGTACTGTCCGAAGGCGGTGCGCAAGCAGCTGGG 1140
DB 1081 AAGGCTTTGCAAGCAGCAGAGGCGGTACTGTCCGAAGGCGGTGCGCAAGCAGCTGGG 1140

```

[illegible]



```

CDS
/gene="Rv2738c"
/complement(3247..3453)
/note="Rv2738c"
N-terminus is highly similar to the N-terminus of the
upstream ORF RV002.07c (78.4% identity in 37 aa overlap);
also similar to AL020958|SC4H8.5 Streptomyces coelicolor
cosmid 4H8 (64aa) opt: 185 z-score: 283.5 E(): 2.9e-08;
39.7% identity in 63 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2738c"
/protein_id="CA15534.1"
/db_xref="GI:2624260"
/db_xref="SPTREMBL:O33281"
/translacion="MAGVRLTEFHERALHFGAAYGSSVLDHVLTLGFGDSRAQAQAI
EDGYEPDVRVRLCADFDVPHDRW"
/complement(3464..4630)
/gene="Rv2739c"
/complement(3464..4630)
/note="Rv2739c" (MTV002.04c) len: 388 aa. Probable
transferase, similar to eg. TR-051560 (EMBL:L28170)
Pseudomonas aeruginosa rhamnosyl transferase (426 aa)
fasta scores; opt: 178 z-score: 226.3 E(): 1.9e-05; 25.9%
identity in 425 aa overlap. Equivalent to Mycobacterium
leprae protein MLCB33.02c (392 aa); fasta scores
9P1294723|MLCB33-2 Mycobacterium leprae cosmid B33 opt:
211.2 z-score: 2364.5 E(): 0; 80.9% identity in 388 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2739c"
/protein_id="CA15535.1"
/db_xref="GI:2624261"
/db_xref="SPTREMBL:O33282"
/translacion="MRAVAVAGPPGHSFPAIALCQFRRAADPTLTGVEWLEAAR
AAGIDAVELDLAIDRLDGRARIRRAAAYLVNPRALPELVASDVITACGG
MAAILDIPWELNPHPLIPGSKLPIGSLAAGTIRGLRDTAKMALGRMRAG
LRQRAAVVEIGLPARDPGLRLIATLPALFVRRPMPAAVAVVGLPEPTDRLVA
IPAGTGVVAVVASTALGTGLTEVALQSLTPEPTVSGSRLLVSGADLPPVA
AVAGISQAEILTEADLVICGGHGVAKTLLAIPVAVVVGSGDQDEIARVROGS
AVLRPLTADLVAAVNEVLSPPFRRAARAAASVAGAADPVVCHDALALAC"
/4674..5123
/gene="Rv2740"
/4674..5123
/note="Rv2740" (MTV002.05) len: 149 aa. unknown, similar
to M. leprae hypothetical protein TR-049850 (EMBL:294723)
MLCB33.03 (B2235_F3_140) (178 aa) fasta scores; opt: 498
z-score: 538.9 E(): 7.3e-23; 51.6% identity in 161 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2740"
/protein_id="CA15536.1"
/db_xref="GI:2624262"
/db_xref="SPTREMBL:O33283"
/translacion="MAELTSPETP TTEAIRAVLAFNALQNEDEFDYDALGDL
VYENVSRIIGRRATILBRMGKRGVFEKIRIRIDGAAGAATTEETDALIIPLAY
OFWYCVPEVDGRRITILMRDLYDYDMFKGLRLVALVLPVSLKATL"
/5355..6932
/gene="PE_PGRS"
/5355..6932
/gene="PE_PGRS"
/5355..6932
/note="Rv2741" (MTV002.06) len: 525 aa. Member of M.
tuberculosis PE_PGRS subfamily, similar to many eg.

Query Match 99.9%; Score 1234.4; LB 1; Length 56414;
Best Local Similarity 99.9%; Pred. No. 6.9e-170;
Matches 1235; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCTTGCAAGATTATCGAAGCTTTCTTACACAGCGTACAGTATCGAGAGGGGTAATC 60

```



Query Match 99.7%; Score 1232; DB 5; Length 1235;  
 Best Local Similarity 99.8%; Pred. No. 3.5e-169;  
 Matches 1232; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 ATCTGAGATTAATGAACTTTCTTCTACACTGAAGGTAAGATGAGAGGGTAAATC 60
DB 1 ATCTGAGATTAATGAACTTTCTTCTACACTGAAGGTAAGATGAGAGGGTAAATC 60
QY 61 ATGGGCTGCTGTTTCGAGCAGAGACCAAAACAAAGATTCGGGGTGCATACCCCG 120
DB 61 ATGGGCTGCTGTTTCGAGCAGAGACCAAAACAAAGATTCGGGGTGCATACCCCG 120
QY 121 GCGGCGCTGCGGAACTAACCCGTGTGGCTAGAGGTGCTCATCAGGAGGTGCCGA 180
DB 121 GCGGCGCTGCGGAACTAACCCGTGTGGCTAGAGGTGCTCATCAGGAGGTGCCGA 180
QY 181 GAGGCTGCGCTTACACAGGCGGATTTCAAGGCGGCGGCAAGCTGTCGAC 240
DB 181 GAGGCTGCGCTTACACAGGCGGATTTCAAGGCGGCGGCAAGCTGTCGAC 240
QY 241 GCGGACGAGGTGCGGCGGAGCTGATTTATGCTCAAGGTCAAGAAAGATAGCGGC 300
DB 241 GCGGACGAGGTGCGGCGGAGCTGATTTATGCTCAAGGTCAAGAAAGATAGCGGC 300
QY 301 GAATAGCGGCGCTGCGACACAGGCGAGATCTTGTTCAGCTTTCGATTTGCCGCTCA 360
DB 301 GAATAGCGGCGCTGCGACACAGGCGAGATCTTGTTCAGCTTTCGATTTGCCGCTCA 360
QY 361 CGATCTTGCACGATGCTGTTGATTCGGGCGACAGCTCAATTCGCTTACAGAGCCGTC 420
DB 361 CGATCTTGCACGATGCTGTTGATTCGGGCGACAGCTCAATTCGCTTACAGAGCCGTC 420
QY 421 CAGACCGCGGCGGCGGAGCTACCCCTGTTGCCCGATGAGCGAAGTGGCGGCTGACTC 480
DB 421 CAGACCGCGGCGGCGGAGCTACCCCTGTTGCCCGATGAGCGAAGTGGCGGCTGACTC 480
QY 481 GCGGCGGCGGCGGCGGAGCTACCCCTGTTGCCCGATGAGCGAAGTGGCGGCTGACTC 540
DB 481 GCGGCGGCGGCGGCGGAGCTACCCCTGTTGCCCGATGAGCGAAGTGGCGGCTGACTC 540
QY 541 GCGGCGGCGGCGGCGGAGCTACCCCTGTTGCCCGATGAGCGAAGTGGCGGCTGACTC 600
DB 541 GCGGCGGCGGCGGCGGAGCTACCCCTGTTGCCCGATGAGCGAAGTGGCGGCTGACTC 600
QY 601 TACAACGCGGCGGCGGAGCTACCCCTGTTGCCCGATGAGCGAAGTGGCGGCTGACTC 660
DB 601 TACAACGCGGCGGCGGAGCTACCCCTGTTGCCCGATGAGCGAAGTGGCGGCTGACTC 660
QY 661 ATGCAACAACTTGGGCAACTGAGCGGAGTTCGGGCGGAGTTCGACACTCGCTACTCA 720
DB 661 ATGCAACAACTTGGGCAACTGAGCGGAGTTCGGGCGGAGTTCGACACTCGCTACTCA 720
QY 721 TCGGCTTACGAGCTGAGGAGTTCGACACTCGCTACTCA 780
DB 721 TCGGCTTACGAGCTGAGGAGTTCGACACTCGCTACTCA 780
QY 781 GTGCGAGGCGGCGGAGCTACCCCTGTTGCCCGATGAGCGAAGTGGCGGCTGACTC 840
DB 781 GTGCGAGGCGGCGGAGCTACCCCTGTTGCCCGATGAGCGAAGTGGCGGCTGACTC 840
QY 841 GGTGCGGCTGAGTGTGATGAGCTGAGCGGAGTTCGACACTCGCTACTCA 900
DB 841 GGTGCGGCTGAGTGTGATGAGCTGAGCGGAGTTCGACACTCGCTACTCA 900
QY 901 ACCGCTTACGAGCTGAGGAGTTCGACACTCGCTACTCA 960
DB 901 ACCGCTTACGAGCTGAGGAGTTCGACACTCGCTACTCA 960
QY 961 ATGCGCGCTGAGTGTGATGAGCTGAGCGGAGTTCGACACTCGCTACTCA 1020
DB 961 ATGCGCGCTGAGTGTGATGAGCTGAGCGGAGTTCGACACTCGCTACTCA 1020
QY 1021 GTGCTGAGCTTGGCGGAGCTGAGCGGAGTTCGACACTCGCTACTCA 1080

```

```

DB 1021 GTGCTGAGCTTGGCGGAGCTGAGCGGAGTTCGACACTCGCTACTCA 1080
QY 1081 AAGGCTTTCGAGCAGCAGGAGGCGTACTGTCGAGAGGCGGCGGCGGAGCTGAGC 1140
DB 1081 AAGGCTTTCGAGCAGCAGGAGGCGTACTGTCGAGAGGCGGCGGCGGAGCTGAGC 1140
QY 1141 GTGCGCTTACGAGGCGGCGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGC 1200
DB 1141 GTGCGCTTACGAGGCGGCGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGC 1200
QY 1201 CACAGCTGCGGAGTAAAGGAGGAGTGTGCGC 1235
DB 1201 CACAGCTGCGGAGTAAAGGAGGAGTGTGCGC 1235
RESULT 8
LOCUS A87608 1228 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 5 from Patent WO9836089.
ACCESSION A87608
VERSION A87608.1 GI:5736248
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1228)
AUTHORS Flohe, L. and Singh, M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FEATURES
source Location/Qualifiers
BASE COUNT 236 a 391 c 382 g 219 t
ORIGIN
Query Match 99.2%; Score 1226.4; DB 5; Length 1228;
Best Local Similarity 99.9%; Pred. No. 2.3e-168;
Matches 1227; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATCTGAGATTAATGAACTTTCTTCTACACTGAAGGTAAGATGAGAGGGTAAATC 60
DB 1 ATCTGAGATTAATGAACTTTCTTCTACACTGAAGGTAAGATGAGAGGGTAAATC 60
QY 61 ATGCGGCTGCTGTTTCGAGCAGAGACCAAAACAAAGATTCGGGGTGCATACCCCG 120
DB 61 ATGCGGCTGCTGTTTCGAGCAGAGACCAAAACAAAGATTCGGGGTGCATACCCCG 120
QY 121 GCGGCGCTGCGGAACTAACCCGTGTGGCTAGAGGTGCTCATCAGGAGGTGCCGA 180
DB 121 GCGGCGCTGCGGAACTAACCCGTGTGGCTAGAGGTGCTCATCAGGAGGTGCCGA 180
QY 181 GAGGCTGCGCTTACACAGGCGGATTTCAAGGCGGCGGCAAGCTGTCGAC 240
DB 181 GAGGCTGCGCTTACACAGGCGGATTTCAAGGCGGCGGCAAGCTGTCGAC 240
QY 241 GCGGAGGCTGAGTGTGATGAGCTGAGCGGAGTTCGACACTCGCTACTCA 300
DB 241 GCGGAGGCTGAGTGTGATGAGCTGAGCGGAGTTCGACACTCGCTACTCA 300
QY 301 GAATAGCGGCGCTGCGACACAGGCGAGATCTTGTTCAGCTTTCGATTTGCCGCTCA 360
DB 301 GAATAGCGGCGCTGCGACACAGGCGAGATCTTGTTCAGCTTTCGATTTGCCGCTCA 360
QY 361 CGTGTTCGACGAGTGTGATGAGCTGAGCGGAGTTCGACACTCGCTACTCA 420
DB 361 CGTGTTCGACGAGTGTGATGAGCTGAGCGGAGTTCGACACTCGCTACTCA 420
QY 421 CAGACCGCGGCGGCGGAGCTACCCCTGTTGCCCGATGAGCGAAGTGGCGGCTGACTC 480

```

Db	421	CAGACGGCCGACGGCGCACTACCCCTGCTTGGCCCGGATGAGGAAGTGGCGGTGCATC	480
Qy	481	GCCGCCAGGTTGGCGCTTACCACTGATGCGAACCAGGGGGCCGCGTGTCTATG	540
Db	481	GCCGCCAGGTTGGCGCTTACCACTGATGCGAACCAGGGGGGGCGGCGTGTCTATG	540
Qy	541	GCGGGGGTGGCGGGCGTGAACGGGGCGGAGCGTGTGTATGAGGGCGGACGGCGGC	600
Db	541	GCGGGGGTGGCGGGCGTGAACGGGGCGGAGCGTGTGTATGAGGGCGGACGGCGGC	600
Qy	601	TACAGCCAGCCCCGATCGCCAAACGGGATGGGCGGACCGTTACGGTTCTAGACATCAAC	660
Db	601	TACAGCCAGCCCCGATCGCCAAACGGGATGGGCGGACCGTTACGGTTCTAGACATCAAC	660
Qy	661	ATCGACAACTTGGGCAACTCTGACGGCGAGTTCTGGGGCGGATCCACACATCGCTACTCA	720
Db	661	ATCGACAACTTGGGCAACTCTGACGGCGAGTTCTGGGGCGGATCCACACATCGCTACTCA	720
Qy	721	TGCGGCTACGAGCTCGAGGGTGGCGTCAAAACGTCGCGACCTGGTATTTGGGGCGTCTGT	780
Db	721	TGCGGCTACGAGCTCGAGGGTGGCGTCAAAACGTCGCGACCTGGTATTTGGGGCGTCTGT	780
Qy	781	GTGCGAGGCGCGAAGGACCCAAATTAGTCTGGAATTCTTTGCGCGCATATGAAACA	840
Db	781	GTGCGAGGCGCGAAGGACCCAAATTAGTCTGGAATTCTTTGCGCGCATATGAAACA	840
Qy	841	GCTGCGGTACGTGTGATATAGCCATCGACACGAGGGGGGCTTTGGAAGGCTCACAGCG	900
Db	841	GCTGCGGTACGTGTGATATAGCCATCGACACGAGGGGGGCTTTGGAAGGCTCACAGCG	900
Qy	901	ACCACTACGACACCCGAGCTTGCGCTGTCACAGACACGCTGTTTACTGCGTGGCGAAC	960
Db	901	ACCACTACGACACCCGAGCTTGCGCTGTCACAGACACGCTGTTTACTGCGTGGCGAAC	960
Qy	961	ATGCCCCCTTGCGTGGCGAAGAGGTGACCTACGCCCTACCAAGCGACATGGCGAT	1020
Db	961	ATGCCCCCTTGCGTGGCGAAGAGGTGACCTACGCCCTACCAAGCGACATGGCGAT	1020
Qy	1021	GTGCTCGAGCTTGGCGACATATGGCTGGCGGGCGGCTGCGGTCGATCCGCACTAACCC	1080
Db	1021	GTGCTCGAGCTTGGCGACATATGGCTGGCGGGCGGCTGCGGTCGATCCGCACTAACCC	1080
Qy	1081	AAAGCTTTTCGACGACAGAAAGGGGGCTTACTGTCTCCGAACGGGTGGCCACACGACTGGGG	1140
Db	1081	AAAGCTTTTCGACGACAGAAAGGGGGCTTACTGTCTCCGAACGGGTGGCCACACGACTGGGG	1140
Qy	1141	GTGCGGTACACGAGCCCGGACGGTGTGCTGACTGTGGCGCGCTGTATGCGCGAG	1200
Db	1141	GTGCGGTACACGAGCCCGGACGGTGTGCTGACTGTGGCGCGCTGTATGCGCGAG	1200
Qy	1201	CACAGTTCGGGATTAAGGGAAGCGATGA	1228
Db	1201	CACAGTTCGGGATTAAGGGAAGCGATGA	1228
RESULT	9		
LOCUS	AB9748	1328 bp	DNA
DEFINITION	Sequence 5 from Patent WO9832862.		
ACCESSION	AB9748		PAT
VERSION	AB9748.1	GI:6738282	
KEYWORDS	unidentified.		
SOURCE	unidentified		
ORGANISM	unclassified		
REFERENCE	1 (bases 1 to 1228)		
AUTHORS	Flohe, L. and Singh, M.		
TITLE	L-LANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM		
JOURNAL	Patent: WO 9832862-A 30-JUL-1998;		
FEATURES	FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)		
source	Location/Qualifiers		
	1..1228		"/organism="unidentified"

BASE COUNT	236 a	391 c	382 g	219 t	
ORIGIN	/db_xref="taxon:32644"				
Query Match	99.2%	Score 1226.4	DB 5	Length 1228	
Best Local Similarity	99.9%	Pred. No. 2.3e-168			
Matches 1227	Conservative	0	Mismatches	1	Indels
		0		Gaps	0
QY	1	ATCTGCAGATTAAATCGAACTTCTTTCACACTGAGCGGTACAGCTATTCAGAGGGGTAATC	60		
DB	1	ATCTTTCGCAATTAAATCGAACTTCTTTCATCTGAGAGGTACAGTATTCGAGAGGGGTAATC	60		
QY	61	ATGCCGCTGGTATTCCGACCGAGACCAAAAAACAAGAAATTCGGGGTGGCCATCAACCCG	120		
DB	61	ATGGCGGTGGTATTCCACCGAGACCAAAAAACAAGAAATTCGGGGTGGCCATCAACCCG	120		
QY	121	GCCGGGCTGGCGGAAGTAACCCGTGTGGCCATGAGTGTCTATTCAGGCAAGTGGCCGA	180		
DB	121	GCCGGGCTGGCGGAAGTAACCCGTGTGGCCATGAGTGTCTATTCAGGCAAGTGGCCGA	180		
QY	181	GAGGGCTGGCTATCAACCAACGCGATTTCAAGGGGCGAGCGCCCAATGCTCGGCACC	240		
DB	181	TAAGGTGGCTATCAACCAACGCGATTTCAAGGGGCGAGCGCCCAATGCTCGGCACC	240		
QY	241	GCCGACCAAGTGTGGCGGACCGCTGATTTATTGCTCAAGTCAAAAGACATAGCGGC	300		
DB	241	GCCGACCAAGTGTGGCGGACCGCTGATTTATTGCTCAAGTCAAAAGACATAGCGGC	300		
QY	301	GAAATAGGGCGGCTGGGAGACGAGGCGATCTTGTCAGCTCTTGCAATTTGGCCGGTCA	360		
DB	301	GAAATAGGGCGGCTGGGAGACGAGGCGATCTTGTCAGCTCTTGCAATTTGGCCGGTCA	360		
QY	361	CGTGCTTGCAACGATGCTGTGTGGATTCGGGACCACTCAATTGCTCTAGAGACGCTC	420		
DB	361	CGTGCTTGCAACGATGCTGTGTGGATTCGGGACCACTCAATTGCTCTAGAGACGCTC	420		
QY	421	CAGACCGCGAGCGGCACTACCCCTGCTGGCCCGCATAGGCAAGTGCCTGGTGACCTC	480		
DB	421	CAGACCGCGAGCGGCACTACCCCTGCTGGCCCGCATAGGCAAGTGCCTGGTGACCTC	480		
QY	481	GCCGCCAGGTTGGCGCTTACCACTGATGCGAACCAGAGGGGGCGCGGTGTCTGATG	540		
DB	481	GCCGCCAGGTTGGCGCTTACCACTGATGCGAACCAGAGGGGGCGCGGTGTCTGATG	540		
QY	541	GCGGGGCTGGCGCGGCTCAACCGGCGAGCTGTGTGATGGGCGCGGACCGCGCGC	600		
DB	541	GCGGGGCTGGCGCGGCTCAACCGGCGAGCTGTGTGATGGGCGCGGACCGCGCGC	600		
QY	601	TACACGCAAGCCGCACTCCGCCAACGCGCATGAGCGGCGGACCGCTTACGGTTAGACATCAAC	660		
DB	601	TACACGCAAGCCGCACTCCGCCAACGCGCATGAGCGGCGGACCGCTTACGGTTAGACATCAAC	660		
QY	661	ATCGCAAACTTGGGAACCTCGACGCGCGGAGTTTGGGGCGGATTCACACTGGCTACTCA	720		
DB	661	ATCGCAAACTTGGGAACCTCGACGCGCGGAGTTTGGGGCGGATTCACACTGGCTACTCA	720		
QY	721	TCGGCTACGAGCTCGAGGATGCCGTCAAAAGCTGGACCTGGTGAATTTGGGGCGCTCTG	780		
DB	721	TCGGCTACGAGCTCGAGGATGCCGTCAAAAGCTGGACCTGGTGAATTTGGGGCGCTCTG	780		
QY	781	GTCGCCAGGGCGCAAGGCGACCAATTAAGTGTGCAATTCATCTTGGCGGATATGAAGCA	840		
DB	781	GTCGCCAGGGCGCAAGGCGACCAATTAAGTGTGCAATTCATCTTGGCGGATATGAAGCA	840		
QY	841	GGTGGCTACTGTGATATAGCCATGACCAAGGCGGCTGTTTCGAAGCTTCAGACCG	900		
DB	841	GGTGGCTACTGTGATATAGCCATGACCAAGGCGGCTGTTTCGAAGCTTCAGACCG	900		
QY	901	ACCACTTAGACCAACCGAGCTGGCGGTGACAGACAGCTGTTTACAGCGTGGGGAAC	960		
DB	901	ACCACTTAGACCAACCGAGCTGGCGGTGACAGACAGCTGTTTACAGCGTGGGGAAC	960		

QY 961 ATGCCCGCTCGGTCGCGAAGAGCTGACCTACGCGCTACCAACGCGACGATGCCGTAT 1020  
 DB 961 ATGCCCGCTCGGTCGCGAAGAGCTGACCTACGCGCTACCAACGCGACGATGCCGTAT 1020  
 QY 1021 GTGCTCGAGCTTCCGACCATGAGTGGCGGGCGGCTGCCGCTCGATCCGCGACTAGCC 1080  
 DB 1021 GTGCTCGAGCTTCCGACCATGAGTGGCGGGCGGCTGCCGCTCGATCCGCGACTAGCC 1080  
 QY 1081 AAGGCTTTCGACGCGACGAGGGGGCTGCTGCTGCAACCTGCGCGACGCGACCTGGGG 1140  
 DB 1081 AAGGCTTTCGACGCGACGAGGGGGCTGCTGCTGCAACGCGCGACGCGACCTGGGG 1140  
 QY 1141 GTGCGCTTCACGAGCGCCCGACAGCTGCTGCGCTGACTCTCGCGCGCTTACCGCGAG 1200  
 DB 1141 GTGCGCTTCACGAGCGCCCGACAGCTGCTGCGCTGACTCTCGCGCGCTTACCGCGAG 1200  
 QY 1201 CACACGTCGGGAGTAAGGAGACGATGA 1228  
 DB 1201 CACACGTCGGGAGTAAGGAGACGATGA 1228

## RESULT 10

A87609 1235 bp DNA PAT 22-JAN-2000  
 LOCUS A87609 Sequence 6 from Patent W09836089.  
 DEFINITION A87609  
 ACCESSION A87609  
 VERSION A87609.1 GI:6736249  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1235)  
 AUTHORS Flohe, L. and Singh, M.  
 TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A-20-AUG-1998:  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES  
 source 1..1235  
 Location/Qualifiers  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 235 a 395 c 384 g 220 t 1 others  
 ORIGIN

Query Match 98.9%; Score 1223; DB 5; Length 1235;  
 Best Local Similarity 99.8%; Pred. No. 7e-168;  
 Matches 1234; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATCTTCAGATTATATCAACTTTCTTACACTGACGATGATGAGAGGGGTATC 60  
 DB 1 ATCTTCAGATTATATCAACTTTCTTACACTGACGATGATGAGAGGGGTATC 60  
 QY 61 ATGCGGTGGGATTCGAGCGGACCAAAACAAGAAATTCGGGTGGCGCATACCCCG 120  
 DB 61 ATGCGGTGGGATTCGAGCGGACCAAAACAAGAAATTCGGGTGGCGCATACCCCG 120  
 QY 121 GCGGCGTGGCGGAATCAACCGTCTGCGCATGAGTCTCATCCAGGACAGTGGCGGA 180  
 DB 121 GCGGCGTGGCGGAATCAACCGTCTGCGCATGAGTCTCATCCAGGACAGTGGCGGA 180  
 QY 181 GAGGCTCGGCTATACCGACGCGGATTTCAAGGGCGGCGGCGCACTGTGGGAC 240  
 DB 181 GAGGCTCGGCTATACCGACGCGGATTTCAAGGGCGGCGGCGCACTGTGGGAC 240  
 QY 241 GCGGACAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
 DB 241 GCGGACAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
 QY 301 GAATACGCGCGCTCGGACGAGGCGGATCTTTCAGCTTCTTGGATTTGGCGGCTCA 360  
 DB 301 GAATACGCGCGCTCGGACGAGGCGGCGGATCTTTCAGCTTCTTGGATTTGGCGGCTCA 360  
 QY 361 CGTGCTTGACCGAGTGGTGTGGATTCGGGCGGCGGCGGCGGCGGCGGCGGCT 420

DB 360 CGTGCTTGACCGAGTGGTGTGGATTCGGGCGGCGGCGGCGGCGGCGGCGGCGGCT 419  
 QY 421 CAGACCGCGGACGGCGCACTACCCCTGCTTCCCGGATGAGCGAAGTCCGCTGCACTC 480  
 DB 420 CAGACCGCGGACGGCGCACTACCCCTGCTTCCCGGATGAGCGAAGTCCGCTGCACTC 479  
 QY 481 GCGGCGGAGTGGGCGCTTACCACTGATGGAACCCCAAGGGGGCGGCGGTGCTCATG 540  
 DB 480 GCGGCGGAGTGGGCGCTTACCACTGATGGAACCCCAAGGGGGCGGCGGTGCTCATG 539  
 QY 541 GCGGCGGAGTGGGCGCTGGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
 DB 540 GCGGCGGAGTGGGCGCTGGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 599  
 QY 601 TACAACGACGCGCGCATGCGCAACGCGCATGGGCGGCGGCGGCGGCGGCGGCGG 660  
 DB 600 TACAACGACGCGCGCATGCGCAACGCGCATGGGCGGCGGCGGCGGCGGCGGCGG 659  
 QY 661 ATGCAAACTTCGCGCACTGACGCGCGGATTCGGGCGGCGGCGGCGGCGGCGGCT 720  
 DB 660 ATGCAAACTTCGCGCACTGACGCGCGGATTCGGGCGGCGGCGGCGGCGGCGGCT 719  
 QY 721 TCGGCTTACGAGCTTCGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 780  
 DB 720 TCGGCTTACGAGCTTCGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 779  
 QY 781 GTGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840  
 DB 780 GTGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 839  
 QY 841 GTGCGGAGCTTGTGATATAGCCATCGACGAGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
 DB 840 GTGCGGAGCTTGTGATATAGCCATCGACGAGGCGGCGGCGGCGGCGGCGGCGGCGG 899  
 QY 901 ACCACTTACGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960  
 DB 900 ACCACTTACGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 959  
 QY 961 ATGCGGCGCTGCGTCCGCAAGACGCTGACGCTACGCGCTGACCAACGCGACATCCG 1020  
 DB 960 ATGCGGCGCTGCGTCCGCAAGACGCTGACGCTACGCGCTGACCAACGCGACATCCG 1019  
 QY 1021 GTGCTCGAGCTTCCGCGACATGCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1080  
 DB 1020 GTGCTCGAGCTTCCGCGACATGCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1079  
 QY 1081 AAGGCTTTTCGACCGCAAAATGCACTTACTGTCGGAACGGGTGGCGGCGGCGG 1140  
 DB 1080 AAGGCTTTTCGACCGCAAAATGCACTTACTGTCGGAACGGGTGGCGGCGGCGG 1139  
 QY 1141 GTGCGCTTACCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
 DB 1140 GTGCGCTTACCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1199  
 QY 1201 CACACGTCGGGAGTAAGGAGACGATGATGTCGCGC 1236  
 DB 1200 CACACGTCGGGAGTAAGGAGACGATGATGTCGCGC 1235

## RESULT 11

A89749 1235 bp DNA PAT 22-JAN-2000  
 LOCUS A89749 Sequence 6 from Patent W09832862.  
 DEFINITION A89749  
 ACCESSION A89749  
 VERSION A89749.1 GI:6738283  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1235)  
 AUTHORS Flohe, L. and Singh, M.  
 TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM

JOURNAL Patent: WO 9832862-A 30-JUL-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)  
 FEATURES Location/Qualifiers  
 source 1.1235  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 235 a 395 c 384 g 220 t 1 others  
 ORIGIN

Query Match 98.9%; Score 1223; DB 5; Length 1235;  
 Best Local Similarity 99.8%; Pred. No. 7e-168;  
 Matches 1234; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```

OY 1 ATCTTGACATTAATGAGACTTCTTACACATGAAAGCTATGAGAGAGGATATC 60
DB 1 ATCTTGACATTAATGAGACTTCTTACACATGAAAGCTATGAGAGAGGATATC 60
OY 61 ATGCGGCTGCTATTCGACCGAGACCAAAACAAAGAAATTCGGGTGGCATCACCCG 120
DB 61 ATGCGGCTGCTATTCGACCGAGACCAAAACAAAGAAATTCGGGTGGCATCACCCG 120
OY 121 GCGGCGTCCGGAAGTAAACCCGCTGCTGAGTCTATCCAGCAGAGTCCGGA 180
DB 121 GCGGCGTCCGGAAGTAAACCCGCTGCTGAGTCTATCCAGCAGAGTCCGGA 180
OY 121 GCGGCGTCCGGAAGTAAACCCGCTGCTGAGTCTATCCAGCAGAGTCCGGA 180
DB 121 GCGGCGTCCGGAAGTAAACCCGCTGCTGAGTCTATCCAGCAGAGTCCGGA 180
OY 181 GAGGGCTGCTATCAGCAGCGGATTCAGAGCGGCGGCGGAGCTGGTGGCACC 240
DB 181 GAGGGCTGCTATCAGCAGCGGATTCAGAGCGGCGGCGGAGCTGGTGGCACC 240
OY 241 GCGGACAGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
DB 241 GCGGACAGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
OY 301 GAATACGCGCGCTGCGGACGAGATTTATGCTCAAGTCAAAAGAACGATAGCGGG 360
DB 301 GAATACGCGCGCTGCGGACGAGATTTATGCTCAAGTCAAAAGAACGATAGCGGG 360
OY 361 GCGGCTGCGGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
DB 361 GCGGCTGCGGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
OY 421 GAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
DB 421 GAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
OY 481 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
DB 481 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
OY 541 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
DB 541 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
OY 601 TACAACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
DB 601 TACAACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
OY 661 ATGCGGCTGCTATTCGACCGAGACCAAAACAAAGAAATTCGGGTGGCATCACCCG 720
DB 661 ATGCGGCTGCTATTCGACCGAGACCAAAACAAAGAAATTCGGGTGGCATCACCCG 720
OY 721 TCGGCGTATGAGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
DB 721 TCGGCGTATGAGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
OY 781 GTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
DB 781 GTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
OY 841 GGTGCGTATGAGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
DB 841 GGTGCGTATGAGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900

```

```

OY 901 ACCACTAGACACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
DB 901 ACCACTAGACACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
OY 961 ATGCGGCTGCTATTCGACCGAGACCAAAACAAAGAAATTCGGGTGGCATCACCCG 1020
DB 961 ATGCGGCTGCTATTCGACCGAGACCAAAACAAAGAAATTCGGGTGGCATCACCCG 1020
OY 1021 GTGCTGAGCTTCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
DB 1021 GTGCTGAGCTTCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
OY 1081 AAGGCTTTCGACCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
DB 1081 AAGGCTTTCGACCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
OY 1141 GTGCGGCTGCTATTCGACCGAGACCAAAACAAAGAAATTCGGGTGGCATCACCCG 1200
DB 1141 GTGCGGCTGCTATTCGACCGAGACCAAAACAAAGAAATTCGGGTGGCATCACCCG 1200
OY 1201 CACAGCTGCGGAGTAAGGAGAGCATGATGCGGCC 1236
DB 1201 CACAGCTGCGGAGTAAGGAGAGCATGATGCGGCC 1235

```

## RESULT 12

LOCUS A87611 1235 bp DNA PAT 22-JAN-2000  
 DEFINITION Sequence 8 from Patent WO9836089.  
 ACCESSION A8 611  
 VERSION A8 311.1 GI:6736251  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 unclassified.  
 unclassified.  
 unclassified.

REFERENCE 1 (bases 1 to 1235)  
 FLOHE, L. and Singh, M.  
 TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE  
 JOURNAL Patent: WO 9836089-A 20-AUG-1998;  
 FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)

FEATURES  
 source Location/Qualifiers  
 1.1235  
 /organism="unidentified"  
 /db\_xref="taxon:32644"

BASE COUNT 235 a 394 c 385 g 220 t  
 ORIGIN

Query Match 98.9%; Score 1222.4; DB 5; Length 1235;  
 Best Local Similarity 99.8%; Pred. No. 8.6e-168;  
 Matches 1234; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```

OY 1 ATCTTGACATTAATGAGACTTCTTACACATGAAAGCTATGAGAGGATATC 60
DB 1 ATCTTGACATTAATGAGACTTCTTACACATGAAAGCTATGAGAGGATATC 60
OY 61 ATGCGGCTGCTATTCGACCGAGACCAAAACAAAGAAATTCGGGTGGCATCACCCG 120
DB 61 ATGCGGCTGCTATTCGACCGAGACCAAAACAAAGAAATTCGGGTGGCATCACCCG 120
OY 121 GCGGCGTCCGGAAGTAAACCCGCTGCTGAGTCTATCCAGCAGAGTCCGGA 180
DB 121 GCGGCGTCCGGAAGTAAACCCGCTGCTGAGTCTATCCAGCAGAGTCCGGA 180
OY 181 GAGGGCTGCTATCAGCAGCGGATTCAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
DB 181 GAGGGCTGCTATCAGCAGCGGATTCAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
OY 241 GCGGACAGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
DB 241 GCGGACAGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300

```



```

OY 301 GAATACGGCCGCTGCGACAGCGGAGATCTGTACAGTCTTGTGATTTGGCCGCTCA 360
DB 301 GATATACGGCCGCTGCGACAGCGGCG-GATCTTGTACAGTCTTGTGATTTGGCCGCTCA 359
OY 361 CGTCTTGACAGCGATGCGTGTGTGATTCGGGACACAGCTCAATTGCGCTACAGACCGCTC 420
DB 361 CGTCTTGACAGCGATGCGTGTGTGATTCGGGACACAGCTCAATTGCGCTACAGACCGCTC 419
OY 421 CAGACCGCGGAGCGGAGATACCCCTGCTTGGCCCGGAGAGGAATGGCCGCTGACATC 480
DB 421 CAGACCGCGGAGCGGAGATACCCCTGCTTGGCCCGGAGAGGAATGGCCGCTGACATC 479
OY 481 GCGGCCAGAGTGGCGCTTACCACTGATGCAACCAAGGAGCGCGGTGTGTGATG 540
DB 481 GCGGCCAGAGTGGCGCTTACCACTGATGCAACCAAGGAGCGCGGTGTGTGATG 539
OY 541 GCGGGGAGTGGCGGCTGCAACCGGCGAGCTGTGTGATGAGGCGCGGACCGCGCG 600
DB 541 GCGGGGAGTGGCGGCTGCAACCGGCGAGCTGTGTGATGAGGCGCGGACCGCGCG 599
OY 601 TACAACGCGAGCGGAGTGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 660
DB 601 TACAACGCGAGCGGAGTGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 659
OY 661 ATGCAACAACTTGGCACTGCAAGCGGAGTGTGCGGCGGAGTGTGATGATCA 720
DB 661 ATGCAACAACTTGGCACTGCAAGCGGAGTGTGCGGCGGAGTGTGATGATCA 719
OY 721 TCGGCTTACGAGCTGAGAGGTGCGCTCAAAAGTGTGCGGAGCTGTGATGATG 780
DB 721 TCGGCTTACGAGCTGAGAGGTGCGCTCAAAAGTGTGCGGAGCTGTGATGATG 779
OY 781 GTCGCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 840
DB 781 GTCGCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 839
OY 841 GTCGCGAGCTGAGTGTGATGATGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 900
DB 841 GTCGCGAGCTGAGTGTGATGATGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 899
OY 901 ACCACTTACGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 960
DB 901 ACCACTTACGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 959
OY 961 ATGCGCGGCTGCGTGGCGGAGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1020
DB 961 ATGCGCGGCTGCGTGGCGGAGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1019
OY 1021 GTCGCTGAGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 1080
DB 1021 GTCGCTGAGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 1079
OY 1081 AAGAGCTTTCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1140
DB 1081 AAGAGCTTTCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1139
OY 1141 GTCGCTTACGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 1200
DB 1141 GTCGCTTACGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 1199
OY 1201 CACACGTCGCGGAGTAAAGGAGGAGGATGATGCGGCC 1236
DB 1201 CACACGTCGCGGAGTAAAGGAGGAGGATGATGCGGCC 1235

```

```

RESULT 13
LOCUS A89751 1235 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 8 from Patent WO9832862.
ACCESSION A89751
VERSION A89751.1 GI:6738285
KEYWORDS
SOURCE unidentified.

```

```

ORGANISM unidentified
REFERENCE 1 (bases 1 to 1235)
AUTHORS Flohe, L. and Singh, M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FEATURES
location/Qualifiers
source 1..1235
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 394 c 385 g 220 t
ORIGIN
Query Match 98.9%; Score 1222.4; DB 5; Length 1235;
Best Local Similarity 99.8%; Pred. No. 8,6e-168;
Matches 1234; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
OY 1 ATCTTGACAGATTATGCACTTCTTCCAGTGAAGCGTACAGTATGAGAGGGTAAATC 60
DB 1 ATCTTGACAGATTATGCACTTCTTCCAGTGAAGCGTACAGTATGAGAGGGTAAATC 60
OY 61 ATGCGCGTGGTATTCGAGCGGAGCAAAACAAAGATTCGGGTGGCCATACCCCG 120
DB 61 ATGCGCGTGGTATTCGAGCGGAGCAAAACAAAGATTCGGGTGGCCATACCCCG 120
OY 121 GCGGCGTGGGAGGAACTAACCGCTGTGCGATAGAGTGTATCCAGAGGAGGAGGAGG 180
DB 121 GCGGCGTGGGAGGAACTAACCGCTGTGCGATAGAGTGTATCCAGAGGAGGAGGAGG 180
OY 181 GAGGCGTGGGATACCGGAGCGGAGTTCAGGCGGAGGCGGCAACTGTGTGGGAGC 240
DB 181 GAGGCGTGGGATACCGGAGCGGAGTTCAGGCGGAGGCGGCAACTGTGTGGGAGC 240
OY 241 GCGGACAGGTGTGGGCGGAGCGTGTATTTGCTCAAGTTCAGGAGGAGGAGGAGG 300
DB 241 GCGGACAGGTGTGGGCGGAGCGTGTATTTGCTCAAGTTCAGGAGGAGGAGGAGG 300
OY 301 GAATACGGCCGCTGCGACAGCGGAGATCTTGTACAGTCTTGTGATTTGGCCGCTCA 360
DB 301 GAATACGGCCGCTGCGACAGCGGAGATCTTGTACAGTCTTGTGATTTGGCCGCTCA 359
OY 361 CGTCTTGACAGCGATGCGTGTGTGATTCGGGACACAGCTCAATTGCGCTACAGACCGCTC 420
DB 361 CGTCTTGACAGCGATGCGTGTGTGATTCGGGACACAGCTCAATTGCGCTACAGACCGCTC 419
OY 421 CAGACCGCGGAGCGGAGCTGCTTGTGATTCGGGACACAGCTCAATTGCGCTACAGACCGCTC 480
DB 421 CAGACCGCGGAGCGGAGCTGCTTGTGATTCGGGACACAGCTCAATTGCGCTACAGACCGCTC 479
OY 481 GCGGCCAGGTTGGGCTTACCACTGATGCAACCAAGGAGGAGGAGGAGGAGGAGGAGG 540
DB 481 GCGGCCAGGTTGGGCTTACCACTGATGCAACCAAGGAGGAGGAGGAGGAGGAGGAGG 539
OY 541 GCGGGGAGTGGCGGCTGCAACCGGCGAGCTGTGATGATGAGGAGGAGGAGGAGGAGG 600
DB 541 GCGGGGAGTGGCGGCTGCAACCGGCGAGCTGTGATGATGAGGAGGAGGAGGAGGAGG 599
OY 601 TACAACGCGAGCGGAGTGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 660
DB 601 TACAACGCGAGCGGAGTGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 659
OY 661 ATGCAACAACTTGGCACTGCAAGCGGAGTGTGCGGCGGAGTGTGATGATCA 720
DB 661 ATGCAACAACTTGGCACTGCAAGCGGAGTGTGCGGCGGAGTGTGATGATCA 719
OY 721 TCGGCTTACGAGCTGAGAGGTGCGCTCAAAAGTGTGCGGAGCTGTGATGATG 780
DB 721 TCGGCTTACGAGCTGAGAGGTGCGCTCAAAAGTGTGCGGAGCTGTGATGATG 779
OY 781 GTCGCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 840
DB 781 GTCGCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 839

```



```

Db 780 GTGCCAGGCGCCAAAGCACCACCAATTAAGTCTGCAATTCAGTGTGCGCATATGAACCA 839
Qy 841 GGTGGGTAAGTGTGATATAGCATACGACGAGGCGGCTGTTTGAAAGCTCAGACCG 900
Db 840 GGTGGGTAAGTGTGATATAGCATACGACGAGGCGGCTGTTTGAAAGCTCAGACCG 899
Qy 901 ACCACCTACGACACCGACGCTGCGCGCTGCGACGACGACGCTTTTACGCGCTGGCGAAC 960
Db 900 ACCACCTACGACACCGACGCTGCGCGCTGCGACGACGACGCTTTTACGCGCTGGCGAAC 959
Qy 961 ATGCCGCGCTGCTGCGCGCAAGAGCTGACCTACGCGCTGACCAACGCGACGATGCGCTAT 1020
Db 960 ATGCCGCGCTGCTGCGCGCAAGAGCTGACCTACGCGCTGACCAACGCGACGATGCGCTAT 1019
Qy 1021 GTGCTGACCTTGCCGACCATGCTGCGCGCGCGCTGCGCTGCAATCGGCTACGCTAGCC 1080
Db 1020 GTGCTGACCTTGCCGACCATGCTGCGCGCGCGCTGCGCTGCAATCGGCTACGCTAGCC 1079
Qy 1081 AAGGCTCTTGCACGACGACGAGGGGGCTACTCTCTCGAAGCGGGTGGCGACGCTGGGG 1140
Db 1080 AAGGCTCTTGCACGACGACGAGGGGGCTACTCTCTCGAAGCGGGTGGCGACGCTGGGG 1139
Qy 1141 GTGCGCTTACCGAGCGCGCGACGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db 1140 GTGCGCTTACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1199
Qy 1201 CACACGTCGGAGTAAGGAAAGGATGTCGGCC 1236
Db 1200 CACACGTCGGAGTAAGGAAAGGATGTCGGCC 1235

RESULT 14
LOCUS A87605 1245 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 2 from Patent WO9836089.
ACCESSION A87605
VERSION A87605.1 GI:6736245
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1245)
AUTHORS Flohe, L. and Singh, M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES
source 1..1245
location/Qualifiers
BASE COUNT 238 a 398 c 387 g 222 t
ORIGIN
Query Match 98.7%; Score 1220; DB 5; Length 1245;
Best Local Similarity 99.5%; Pred. No. 1.9e-167;
Matches 1236; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

```

```

Qy 235 GGCACCGCCGACAGGCTGTGGGCGGACGCTGATTTATGCTCAAGGCTCAAGAACGATA 294
Db 241 GGCACCGCCGACAGGCTGTGGGCGGACGCTGATTTATGCTCAAGGCTCAAGAACGATA 300
Qy 295 GCGCGGAATACGCGCGCTGCGACGAGGCGAGATCTGTTACGCTTCTGCAATTTGGCC 354
Db 301 GCGCGGAATACGCGCGCTGCGACGAGGCGAGATCTGTTACGCTTCTGCAATTTGGCC 360
Qy 355 GCGTACGCTTGTGACGAGATGCGTTGTTGATTCGCGACACGCTAATTCCTACGAG 414
Db 361 GCGTACGCTTGTGACGAGATGCGTTGTTGATTCGCGACACGCTAATTCCTACGAG 420
Qy 415 ACCGTCCAGACCGCGACGAGGCGCACTACCTGCTGCTGCGCGGATGAGGAAAGTGGCGGT 474
Db 421 ACCGTCCAGACCGCGACGAGGCGCACTACCTGCTGCTGCGCGGATGAGGAAAGTGGCGGT 480
Qy 475 CGACTGCGCGCCAGGTTGGCGCTTACCACTGATGCGAACCAGAGGAGGCGCGGCTGTG 534
Db 481 CGACTGCGCGCCAGGTTGGCGCTTACCACTGATGCGAACCAGAGGAGGCGCGGCTGTG 540
Qy 535 CTGATGCGCGGCTGCGCGCGCTGCGACCGCGCGACGCTGCTGCTGCTGCTGCTGCTGCT 594
Db 541 CTGATGCGCGGCTGCGCGCGCTGCGACCGCGCGACGCTGCTGCTGCTGCTGCTGCTGCT 600
Qy 595 GCGGCTTCAACGCGCGCGCGATGCGCAAGGAGATGGGCGGACGCTTACGCTTACGAC 654
Db 601 GCGGCTTCAACGCGCGCGCGATGCGCAAGGAGATGGGCGGACGCTTACGCTTACGAC 660
Qy 655 ATCAACATGCAACAACTTGGCACTGCGCACTGCGCGCGGATCTGCGCGGATCAGACTGCG 714
Db 661 ATCAACATGCAACAACTTGGCACTGCGCACTGCGCGCGGATCTGCGCGGATCAGACTGCG 720
Qy 715 TACTCATGCGCTTACGAGCTGAGGCTGCGCTCAACGCTGCTGCTGCTGCTGCTGCTGCT 774
Db 721 TACTCATGCGCTTACGAGCTGAGGCTGCGCTCAACGCTGCTGCTGCTGCTGCTGCTGCT 780
Qy 775 GTCTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 834
Db 781 GTCTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
Qy 835 AAACGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 894
Db 841 AAACGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Qy 895 CGACGCGACCTTACGACACGCTGCGACGCTGCGACGCTGCTGCTGCTGCTGCTGCTGCTG 954
Db 901 CGACGCGACCTTACGACACGCTGCGACGCTGCGACGCTGCTGCTGCTGCTGCTGCTGCTG 960
Qy 955 GCGAATATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1014
Db 961 GCGAATATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
Qy 1015 CCGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1074
Db 1021 CCGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Qy 1075 TACCCAAAGGCTTTCGACGACGAGAGGCGCTTACTGCTCGAAGCGGTGGCCACCGAC 1134
Db 1081 TACCCAAAGGCTTTCGACGACGAGAGGCGCTTACTGCTCGAAGCGGTGGCCACCGAC 1140
Qy 1135 CTGGGGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1194
Db 1141 CTGGGGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
Qy 1195 GCGGACGACGCTGCGGAGTAAGGAAAGGATGATGTCGGCC 1236
Db 1201 GCGGACGACGCTGCGGAGTAAGGAAAGGATGATGTCGGCC 1242

RESULT 15
LOCUS A89745 1245 bp DNA PAT 22-JAN-2000

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:11:43 ; Search time 458.59 Seconds  
(without alignment)

674.322 Million cell updates/sec

Title: US-09-362-485-10

Perfect score: 1236  
Sequence: 1 ATCTTCGAGATTATCGAAC.....GGGAAGCGATGATGTCGCC 1236

Scoring table: IDENTITY\_NDC

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

```
Minimum DB seq length: 0
Maximum DB seq length: 1000000
```

Maximum DB seq length: 1000000

Post-processing: Minimum Match 08

Listing first 45 summaries

Database : N\_Geneseq\_36: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1220	98.7	1245	1	V49666	Mycobacterium tube
2	1220	98.7	1260	1	V49510	Mycobacterium sp.
3	1220	98.7	1260	1	V49625	Mycobacterium tube
4	467.6	37.8	682	1	V49511	Mycobacterium marit
5	277.2	22.4	1125	1	N91453	Sequence of heat-r
6	1212.4	17.2	28171	1	V52155	Streptococcus pneu
7	178.4	14.4	9280	1	V74442	Staphylococcus aut
8	98.8	7.7	1074	1	T67971	H. pylori membrane
9	69.2	5.6	544	1	V43039	Streptococcus pneu
10	57.6	4.7	31096	1	V74370	Staphylococcus aut
11	50.8	4.1	390	1	Q21833	Randomising oligon
12	50.8	4.1	360	1	Q36859	PCR primer for 5'
13	50.4	4.1	1413	1	X34205	Mycobacterium spec
14	50.4	4.1	1722	1	X34206	Mycobacterium spec
15	50.4	4.1	15872	1	T68715	Streptomyces venez
16	49	4.0	1269	1	Q00305	Sequence encoding
17	47.6	3.9	3946	1	T93610	Mycobacterium tube
18	47.6	3.9	28596	1	T06769	Sorangium cellulos
19	47.6	3.9	28958	1	T89956	Sorangium cellulos
20	47.6	3.9	49377	1	V05287	The soraphin bios
21	46.4	3.8	985	1	V45439	Mycobacterium tube
22	46.4	3.8	985	1	V64548	M. tuberculosis im
23	46	3.7	1620	1	Q22482	groEL-1 gene codit
24	46	3.7	2582	1	T73117	Actinoplanes sp. a
25	46	3.7	2668	1	Q22485	groEL-1 gene. Rec
26	46	3.7	17955	1	V56642	Actinoplanes sp. a
27	45.4	3.7	833	1	Q64203	snab gene encoding
28	45.4	3.7	5392	1	Q64201	Sequence compisint
29	42.4	3.4	1833	1	Q64206	snR gene encoding
30	42.4	3.4	2185	1	V84066	Clone P5-4 encodin
31	42.4	3.4	2185	1	Q99365	S. lividans proteas
32	42.4	3.4	24379	1	T93095	Streptomyces frenc
33	42.4	3.4	24379	1	T52925	Streptomyces roseo
34	42.2	3.4	329	1	V44425	Mycobacterium tube

C 35	42.2	3.4	339.1	V6453.4	M. tuberculosis im
C 36	4.2	3.4	396.5	T9361.0	Mycobacterium tube
C 37	4.2	3.4	401.8	O6387.9	Polyhydroxyalkanoa
C 38	41.8	3.4	882.1	V4440.3	Mycobacterium tube
C 39	41.8	3.4	882.1	V6451.2	M. tuberculosis im
C 40	41.6	3.4	1203.6	O0465.8	FHA structural gen
C 41	41.2	3.3	241.4	O0532.6	Sequence encoding
C 42	4.1	3.3	2987.9	O4680.6	eryA region of S.
C 43	40.8	3.3	215.1	O7625.2	HSV-2 protease, IC
C 44	40.8	3.3	215.1	O7626.1	HSV-2 protease/ICP
C 45	40.8	3.3	247.2	O8457.1	HSV-2 UL26 gene. N

## ALIGNMENTS

RESULT	1
ID	V49626
DT	V49626 standard; DNA; 1245 BP.
AC	V49626;
DT	20-NOV-1998 (first entry)
DE	Mycobacterium tuberculosis L-Alanine dehydrogenase 40 kD antigen.
KM	Str: Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS	Mycobacterium tuberculosis.
PN	M09836089-42.
PD	20-AUG-1998.
PF	29-JAN-1998; E00483.
PR	29-JAN-1997; E-101338.
PA	(FLOH/) FLOHE L.
PI	Flohe L, Hutter B, Kolb A, Singh M;
DR	WPI, 98-457123/39.
PT	Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT	- useful for, e.g. for diagnosis, differentiation of strains,
PT	monitoring vaccination and identification of mycobacterial
PT	inhibitors
PS	Disclosure; fig 3. 19; 55pp; German.
CC	The Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC	the production of kits for diagnosing tuberculosis (TB) and other
CC	mycobacterial infections in humans or animals. Kits are used for direct
CC	diagnosis of TB on clinical samples (e.g. body fluids) and can
CC	differentiate between pathogenic and non-virulent strains, e.g. for
CC	identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC	also be used to identify substances that inhibit mycobacteria, for
CC	combating epidemics and for vaccination follow-up. Oligonucleotides
CC	derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC	also for culture confirmation of isolated strains and for chromosome
CC	fingerprinting to detect/differentiate between mycobacteria, and for
CC	L-alanine-specific biotransformation reactions. AlaDH is an early
CC	antigen, present extracellularly after only a few days of growth, making
CC	it an ideal drug target.
SO	Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;
Query Match	98.7%; Score 1220; DB 1; Length 1245;
Best Local Similarity	99.5%; Ped. No. 3.4e-245;
Matches 1235; Conservative	0; Mismatches 0; Indels 6; Gaps 1
QY	1 ATCTGCGATTAATGAACTTCTCTCACTCGAAGCGGTACGATGCGAGAGCGGTATG 60
DB	1 ATCTGCGATTAATGAACTTCTCTCACTCGAAGCGGTACGATGCGAGAGCGGTATG 60
OY	61 ATGGCGTGGTATTCGAGCCGAGACCAAAACAGC-----ATTTCGGGTGGCCATC 114
DB	61 ATGGCGTGGTATTCGAGCCGAGACCAAAACAGC-----ATTTCGGGTGGCCATC 120
OY	115 ACCCGGCGCGGCTCGCGGAATTAACCCGTCGTGGCCATGAGGTCTCATCCAGCAGGT 174
DB	121 ACCCGGCGCGGCTCGCGGAATTAACCCGTCGTGGCCATGAGGTCTCATCCAGCAGGT 180
OY	175 GCGGGAAGGCGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCGCGCAACTGTC 234
DB	181 GCGGGAAGGCGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCGCGCAACTGTC 240

QY 235 GGCACCCGCGACAGTGTGGGCGGACGCTGATTATTGCTCAAGTCAAGAACCGATA 294  
 |||||||  
 Db 241 GGCACCGCGGACAGTGTGGGCGGACGCTGATTATTGCTCAAGTCAAGAACCGATA 300  
 QY 295 GCGGCGGAATACGGCCGCTCGACACGCGGACGATCTTGTTCAGCTTTGCAATTTGGCC 354  
 |||||||  
 Db 301 GCGGCGGAATACGGCCGCTCGACACGCGGACGATCTTGTTCAGCTTTGCAATTTGGCC 360  
 QY 355 GCGTCAGCTGTTCACACGATGCGTGTGATTCCGCGACAGCTCAATTTGCCACGAG 414  
 |||||||  
 Db 361 GCGTCAGCTGTTCACACGATGCGTGTGATTCCGCGACAGCTCAATTTGCCACGAG 420  
 QY 415 ACCGTCCAGACCGCGGACGCGGACGCTGCTTGTCCCGATGAGCGAAGTGGCGGT 474  
 |||||||  
 Db 421 ACCGTCCAGACCGCGGACGCGGACGCTGCTTGTCCCGATGAGCGAAGTGGCGGT 480  
 QY 475 CGACTCGCGCGGACGCTGCGCTTACCACTGATGCAACCCAAAGGGGCGCGGTGTG 534  
 |||||||  
 Db 481 CGACTCGCGCGGACGCTGCGCTTACCACTGATGCAACCCAAAGGGGCGCGGTGTG 540  
 QY 535 CTGATGGGCGGCGGCGGCGGCGTCAACCGCGGACGCTGCTGATGCGCGCGGCGAC 594  
 |||||||  
 Db 541 CTGATGGGCGGCGGCGGCGGCGTCAACCGCGGACGCTGCTGATGCGCGCGGCGAC 600  
 QY 595 GCGGCGTCAACGAGCGCGCATGCGCAAGCGCATGCGCGGACCGCTTACGCTTACAG 654  
 |||||||  
 Db 601 GCGGCGTCAACGAGCGCGCATGCGCAAGCGCATGCGCGGACCGCTTACGCTTACAG 660  
 QY 655 ATCAACATCGCAAACTTGGGCACTGAGCGCGGATCTCGCGCGGATCCACACTGCG 714  
 |||||||  
 Db 661 ATCAACATCGCAAACTTGGGCACTGAGCGCGGATCTCGCGCGGATCCACACTGCG 720  
 QY 715 TACTCATCGGCGTCAAGAGCTGAGGCGTCCGTCAAACGCTGCGGACGCTGATTTGGGCG 774  
 |||||||  
 Db 721 TACTCATCGGCGTCAAGAGCTGAGGCGTCCGTCAAACGCTGCGGACGCTGATTTGGGCG 780  
 QY 775 GTCTGTGTCAGAGCGCGCAAGGACCAAAATAGTCTGCAATTCATTTGCGCATATG 834  
 |||||||  
 Db 781 GTCTGTGTCAGAGCGCGCAAGGACCAAAATAGTCTGCAATTCATTTGCGCATATG 840  
 QY 835 AAACGAGTGGGCTGATGCTGATATAGCATCGACGCGGCGGCTGTTGGAAGGCTCA 894  
 |||||||  
 Db 841 AAACGAGTGGGCTGATGCTGATATAGCATCGACGCGGCGGCTGTTGGAAGGCTCA 900  
 QY 895 CGACCGACCACTAGACGACGCGGACGCTGCGCGGACGACGCTGTTTACTGCGTG 954  
 |||||||  
 Db 901 CGACCGACCACTAGACGACGCGGACGCTGCGCGGACGACGCTGTTTACTGCGTG 960  
 QY 955 GCGAATCATGCGCGCTCGGTGCGGAAGACGTGACCTACGCGCTGACCAACGCGCATG 1014  
 |||||||  
 Db 961 GCGAATCATGCGCGCTCGGTGCGGAAGACGTGACCTACGCGCTGACCAACGCGCATG 1020  
 QY 1015 CCGATGTCGTCGAGCTTCCGACCATGCTGCGGCGGCGGCTGCTGCAATTCGGCA 1074  
 |||||||  
 Db 1021 CCGATGTCGTCGAGCTTCCGACCATGCTGCGGCGGCGGCTGCTGCAATTCGGCA 1080  
 QY 1075 CTAGCGAAGGCTCTTTCAGACGCAAGAGGCGTACTGTCGGAAGGGGTGGCCACGAC 1134  
 |||||||  
 Db 1081 CTAGCGAAGGCTCTTTCAGACGCAAGAGGCGTACTGTCGGAAGGGGTGGCCACGAC 1140  
 QY 1135 CTGGGGGTGCGCTTACCGAGCGCGGACGCTGCGCTGACTCTGCGCGCTGTTAC 1194  
 |||||||  
 Db 1141 CTGGGGGTGCGCTTACCGAGCGCGGACGCTGCGCTGACTCTGCGCGCTGTTAC 1200  
 QY 1195 GCGGAGCACAGTGGGAGTAAGGGAAGCATGATGGGCG 1236  
 |||||||  
 Db 1201 GCGGAGCACAGTGGGAGTAAGGGAAGCATGATGGGCG 1242

RESULT 2  
 ID V49510 standard: DNA: 1260 BP.  
 AC V49510:

DT 20-OCT-1998 (first entry)  
 DE Mycobacterium sp. AladH DNA.  
 KM Alanine dehydrogenase; AladH; ADH; diagnosis; tuberculosis; pathogen;  
 KM swimmers disease; vaccine; epidemic; infection; identification; ss.  
 OS Mycobacterium sp.  
 PN W09832862-A2.  
 PD 30-JUL-1998.  
 PF 29-JAN-1998: E00484.  
 PR 29-JAN-1997: EP-101339.  
 PA (FLOH/) FLOHE L.  
 PI FLOHE L, Hutter B, Kolik A, Singh M;  
 DR WPI: 98-427958/35.  
 PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum  
 PT - used for diagnosis of tuberculosis and other mycobacterial  
 PT diseases, also for treatment and prevention, for drug screening and  
 PT for bio-transformation  
 PS Disclosure: Page 11: 57pp: German.  
 CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated  
 CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis  
 CC and other mycobacterial infections (including 'swimmers' disease', caused  
 CC by M. marinum, a fish pathogen) in humans or animals. The protein can  
 CC also be used for control of epidemics and for vaccination. To screen for  
 CC agents with anti-mycobacterial activity, and in bio-transformations that  
 CC are specific for L-alanine. Also Mycobacteria can be identified by  
 CC analysis of genomic ADH sequences. ADH is an antigen thpt is secreted  
 CC early during infection.  
 SQ Sequence 1260 BP: 243 A; 403 C; 389 G; 225 T;

Query Match 98.7%; Score 1220; DB 1; Length 1260;  
 Best Local Similarity 99.5%; Pred. No. 3,4e-245;  
 Matches 1236; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

QY 1 ATCTTGAGATTAATGCACTTCTTCACTGACGCTACGATACGAGGCGTAATC 60  
 Db 16 ATCTTGAGATTAATGCACTTCTTCACTGACGCTACGATACGAGGCGTAATC 75  
 QY 61 ATGCGCGTCTGATTCGCGACGAGACCAAAACACG-----AATTCGCGGTGGCCATC 114  
 |||||||  
 Db 76 ATGCGCGTCTGATTCGCGACGAGACCAAAACACGATTCGCGGTGGCCATC 135  
 QY 115 ACCCGGCGCGGCGTGGCGAATACCGCTGCGCATGAGGTGCTCATCGAGGAGT 174  
 |||||||  
 Db 136 ACCCGGCGCGGCGTGGCGAATACCGCTGCGCATGAGGTGCTCATCGAGGAGT 195  
 QY 175 GCGGAGAGGCGTGGCTATCACCGAGCGGATTTCAAGCGGCGGCGGCACTGCTC 224  
 |||||||  
 Db 196 GCGGAGAGGCGTGGCTATCACCGAGCGGATTTCAAGCGGCGGCGGCACTGCTC 255  
 QY 235 GGCACCCCGGACGAGTGTGGCGGACGCTGATTTATGCTCAAGGCAAGACGATA 294  
 |||||||  
 Db 256 GGCACCCCGGACGAGTGTGGCGGACGCTGATTTATGCTCAAGGCAAGACGATA 315  
 QY 295 GCGGCGAATACGCGCGCTGCGACACGCGGCGAGATCTTGTTCACGTTCTTGCATTGGCC 354  
 |||||||  
 Db 316 GCGGCGAATACGCGCGCTGCGACACGCGGCGAGATCTTGTTCACGTTCTTGCATTGGCC 375  
 QY 355 GCGTCAGTGTTCACACGATGCTGTTGATTTCCGGCACACGATCAATTTGCCCTACGAG 414  
 |||||||  
 Db 376 GCGTCAGTGTTCACACGATGCTGTTGATTTCCGGCACACGATCAATTTGCCCTACGAG 435  
 QY 415 ACCGTCCAGACCGCGGACGCGGACGCTACCGCTGCTTCCCGGATGAGCGAAGTGGCGGT 474  
 |||||||  
 Db 436 ACCGTCCAGACCGCGGACGCGGACGCTACCGCTGCTTCCCGGATGAGCGAAGTGGCGGT 495  
 QY 475 CGACTCGCGCGGACGCTTACCACTGATGGAACCAAGGGGCGCGGCTGTG 534  
 |||||||  
 Db 496 CGACTCGCGCGGACGCTTACCACTGATGGAACCAAGGGGCGCGGCTGTG 555  
 QY 535 CTGATGGGCGGCGGCGGCGGCGTGAACCGGCGGACGCTGCTGATGCGGCGCGGCGAC 594  
 |||||||  
 Db 556 CTGATGGGCGGCGGCGGCGGCGTGAACCGGCGGACGCTGCTGATGCGGCGCGGCGAC 615

CC derived from AlADH are used similarly in diagnostic hybridisation tests, also for culture confirmation of isolated strains and for chromosome fingerprinting to detect/differentiate between mycobacteria, and for CC L-alanine-specific bio-transformation reactions. AlADH is an early CC antigen, present extracellularly after only a few days of growth, making it an ideal drug target.

Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 98.7%; Score 1220; DB 1; Length 1260; Best Local Similarity 99.5%; Pred. No. 3.4e-245; Matches 1236; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

1 ATCTTGACAGATTAAATGACACTTCTTACACTGAAAGCTACGATGTCAGAGGTTATC 60  
 16 ATCTTGACAGATTAAATGACACTTCTTACACTGAAAGCTACGATGTCAGAGGTTATC 75  
 61 ATGGCGGTGGTATTCGCGAGACCAAAACAAACG-----AATTGCGGGTGGCCATC 114  
 76 ATGGCGGTGGTATTCGCGAGACCAAAACAAACGATTTCGATTCGAGGTTATC 135  
 115 ACCCGCGCGCGGTGCGGAACTAACCCGTCGTGACATGAGTGTCTATCCAGGCAAGT 174  
 136 ACCCGCGCGCGGTGCGGAACTAACCCGTCGTGACATGAGTGTCTATCCAGGCAAGT 195  
 175 GCGGAGAGGCTCGGCTATACACGACGCGGATTTCAAGGGGCGAGCGCACTGCTC 234  
 196 GCGGAGAGGCTCGGCTATACACGACGCGGATTTCAAGGGGCGAGCGCACTGCTC 255  
 235 GGCACCGCGGACGAGTGGGCGGAGCGTGAATTAATGCTAAGTCAAGAAACGATA 294  
 256 GGCACCGCGGACGAGTGGGCGGAGCGTGAATTAATGCTAAGTCAAGAAACGATA 315  
 295 GCGGAGGAAATACGCGCGCTGCGACACGCGGAGATCTTGTACGCTTTCGATTGGCC 354  
 316 GCGGAGGAAATACGCGCGCTGCGACACGCGGAGATCTTGTACGCTTTCGATTGGCC 375  
 355 GCGTACAGTGTTCGACCG "GCGTTGTGATTCGCGACCA TCAATTGCTACGAG 414  
 376 GCGTACAGTGTTCGACCG "GCGTTGTGATTCGCGACCA TCAATTGCTACGAG 435  
 415 ACCGTCCAGACCGCGGAGCGGACGACGCGGATGTTGCGCGGATGAGGAAGTCCGGT 474  
 436 ACCGTCCAGACCGCGGAGCGGACGACGCGGATGTTGCGCGGATGAGGAAGTCCGGT 495  
 475 CGACTCGCGCGCGAGTGGCGCTTACCACTGATGCAACCAAGGGCGCGGCTGTG 534  
 496 CGACTCGCGCGCGAGTGGCGCTTACCACTGATGCAACCAAGGGCGCGGCTGTG 555  
 535 CTGATGGGCGGGGTGCGCGCTTACCACTGATGCAACCAAGGGCGCGGCTGTG 594  
 556 CTGATGGGCGGGGTGCGCGCTTACCACTGATGCAACCAAGGGCGCGGCTGTG 615  
 595 GCGGCTTACAGGAGCGCGGCTTACCACTGATGCAACCAAGGGCGCGGCTGTG 654  
 616 GCGGCTTACAGGAGCGCGGCTTACCACTGATGCAACCAAGGGCGCGGCTGTG 675  
 655 ATCAACATGACAAATCTTGCACTGCAAGCGGATTCGCGGCGATCCACTGCG 714  
 676 ATCAACATGACAAATCTTGCACTGCAAGCGGATTCGCGGCGATCCACTGCG 735  
 715 TACTATCGGCTTACGAGTCAAGGTCGCGCAACGTCGCGGATTCGCGGATTC 774  
 736 TACTATCGGCTTACGAGTCAAGGTCGCGCAACGTCGCGGATTCGCGGATTC 795  
 775 GTCTGTGTCAGAGCGCGCAAGCGCAACCAATTAATGCTCAATTCATCTGCGCATATG 834  
 796 GTCTGTGTCAGAGCGCGCAAGCGCAACCAATTAATGCTCAATTCATCTGCGCATATG 855  
 835 AAACCAAGTGGGTACTGTGATATGACATGACGACGAGGCGGCTGTTCCAGAGCTCA 894  
 856 AAACCAAGTGGGTACTGTGATATGACATGACGACGAGGCGGCTGTTCCAGAGCTCA 915

CC combatting epidemics and for vaccination follow-up. Oligonucleotides

CC also for culture confirmation of isolated strains and for chromosome fingerprinting to detect/differentiate between mycobacteria, and for CC L-alanine-specific bio-transformation reactions. AlADH is an early CC antigen, present extracellularly after only a few days of growth, making it an ideal drug target.

Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 98.7%; Score 1220; DB 1; Length 1260; Best Local Similarity 99.5%; Pred. No. 3.4e-245; Matches 1236; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

1 ATCTTGACAGATTAAATGACACTTCTTACACTGAAAGCTACGATGTCAGAGGTTATC 60  
 16 ATCTTGACAGATTAAATGACACTTCTTACACTGAAAGCTACGATGTCAGAGGTTATC 75  
 61 ATGGCGGTGGTATTCGCGAGACCAAAACAAACG-----AATTGCGGGTGGCCATC 114  
 76 ATGGCGGTGGTATTCGCGAGACCAAAACAAACGATTTCGATTCGAGGTTATC 135  
 115 ACCCGCGCGCGGTGCGGAACTAACCCGTCGTGACATGAGTGTCTATCCAGGCAAGT 174  
 136 ACCCGCGCGCGGTGCGGAACTAACCCGTCGTGACATGAGTGTCTATCCAGGCAAGT 195  
 175 GCGGAGAGGCTCGGCTATACACGACGCGGATTTCAAGGGGCGAGCGCACTGCTC 234  
 196 GCGGAGAGGCTCGGCTATACACGACGCGGATTTCAAGGGGCGAGCGCACTGCTC 255  
 235 GGCACCGCGGACGAGTGGGCGGAGCGTGAATTAATGCTAAGTCAAGAAACGATA 294  
 256 GGCACCGCGGACGAGTGGGCGGAGCGTGAATTAATGCTAAGTCAAGAAACGATA 315  
 295 GCGGAGGAAATACGCGCGCTGCGACACGCGGAGATCTTGTACGCTTTCGATTGGCC 354  
 316 GCGGAGGAAATACGCGCGCTGCGACACGCGGAGATCTTGTACGCTTTCGATTGGCC 375  
 355 GCGTACAGTGTTCGACCG "GCGTTGTGATTCGCGACCA TCAATTGCTACGAG 414  
 376 GCGTACAGTGTTCGACCG "GCGTTGTGATTCGCGACCA TCAATTGCTACGAG 435  
 415 ACCGTCCAGACCGCGGAGCGGACGACGCGGATGTTGCGCGGATGAGGAAGTCCGGT 474  
 436 ACCGTCCAGACCGCGGAGCGGACGACGCGGATGTTGCGCGGATGAGGAAGTCCGGT 495  
 475 CGACTCGCGCGCGAGTGGCGCTTACCACTGATGCAACCAAGGGCGCGGCTGTG 534  
 496 CGACTCGCGCGCGAGTGGCGCTTACCACTGATGCAACCAAGGGCGCGGCTGTG 555  
 535 CTGATGGGCGGGGTGCGCGCTTACCACTGATGCAACCAAGGGCGCGGCTGTG 594  
 556 CTGATGGGCGGGGTGCGCGCTTACCACTGATGCAACCAAGGGCGCGGCTGTG 615  
 595 GCGGCTTACAGGAGCGCGGCTTACCACTGATGCAACCAAGGGCGCGGCTGTG 654  
 616 GCGGCTTACAGGAGCGCGGCTTACCACTGATGCAACCAAGGGCGCGGCTGTG 675  
 655 ATCAACATGACAAATCTTGCACTGCAAGCGGATTCGCGGCGATCCACTGCG 714  
 676 ATCAACATGACAAATCTTGCACTGCAAGCGGATTCGCGGCGATCCACTGCG 735  
 715 TACTATCGGCTTACGAGTCAAGGTCGCGCAACGTCGCGGATTCGCGGATTC 774  
 736 TACTATCGGCTTACGAGTCAAGGTCGCGCAACGTCGCGGATTCGCGGATTC 795  
 775 GTCTGTGTCAGAGCGCGCAAGCGCAACCAATTAATGCTCAATTCATCTGCGCATATG 834  
 796 GTCTGTGTCAGAGCGCGCAAGCGCAACCAATTAATGCTCAATTCATCTGCGCATATG 855  
 835 AAACCAAGTGGGTACTGTGATATGACATGACGACGAGGCGGCTGTTCCAGAGCTCA 894  
 856 AAACCAAGTGGGTACTGTGATATGACATGACGACGAGGCGGCTGTTCCAGAGCTCA 915

OY 895 CGACGACACCTAGACACACCGAGCTTCCGCTGACGACGACGCTTTTACTGCTG 954  
 |||||||  
 DB 916 CGACGACACCTAGACACACCGAGCTTCCGCTGACGACGACGCTTTTACTGCTG 975  
 |||||||  
 OY 955 GCGAATATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1014  
 |||||||  
 DB 976 GCGAATATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1035  
 |||||||  
 OY 1015 CGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1074  
 |||||||  
 DB 1036 CGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1095  
 |||||||  
 OY 1075 CTAGCCAAAGGCTCTTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1134  
 |||||||  
 DB 1096 CTAGCCAAAGGCTCTTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1155  
 |||||||  
 OY 1135 CTGGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1194  
 |||||||  
 DB 1156 CTGGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1215  
 |||||||  
 OY 1195 GCGGAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1236  
 |||||||  
 DB 1216 GCGGAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1257  
 |||||||

# RESULT 4

V49511  
 ID V49511 standard; DNA: 682 BP.

AC V49511.  
 DT 20-OCT-1998 (first entry)  
 DE Mycobacterium marinum Mar3 DNA.  
 KM Alanine dehydrogenase; AlADH; ADH; diagnosis; tuberculosis; pathogen;  
 OS Swimmers disease; vaccine; epidemic; infection; identification; ss.  
 KM Mycobacterium marinum.  
 FH Key location/Qualifiers  
 FT CDS 1..682  
 FT /\*tag= a  
 FT /codon\_start= 3  
 FT /product= "Mar3"  
 FT /note= "Alanine dehydrogenase"

PD 30-JUL-1998.  
 PF 29-JAN-1998; E00484.  
 PR 29-JAN-1997; EP-101339.  
 PA (FLOHE/) FLOHE L.  
 PI FLOHE L, Butler B, Kolk A, Singh M;  
 DR MPI: 98-427958/35.  
 DR P-PSDB: W64481.  
 PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum  
 PT - used for diagnosis of tuberculosis and other mycobacterial  
 PT diseases, also for treatment and prevention, for drug screening and  
 PT for bio-transformation  
 PS Claim 1: Page 34-35: 57pp; German.  
 CC This sequence encodes an alanine dehydrogenase (ADH) protein, Mar3  
 CC isolated from Mycobacterium marinum. This protein is used to diagnose  
 CC tuberculosis and other mycobacterial infections (including 'swimmers'  
 CC disease', caused by M. marinum, a fish pathogen) in humans or animals.  
 CC The protein can also be used for control of epidemics and for  
 CC vaccination, to screen for agents with anti-mycobacterial activity, and  
 CC in bio-transformations that are specific for L-alanine. Also mycobacteria  
 CC can be identified by analysis of genomic ADH sequences. ADH is an antigen  
 CC that is secreted early during infection.  
 SQ Sequence 682 BP; 105 A; 254 C; 225 G; 98 T;

Query Match 37.8%; Score 467.6; DB 1: Length 682;  
 Best Local Similarity 80.4%; Pred. No. 5,1e-89;  
 Matches 548; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

OY 98 AATTCCGGTGGCCATCACCCGCGCGCTGCGGAACTAACCCGCTGCGCATGAG 157  
 |||||||  
 DB 1 AATTCCGGTGGCCATCACCCGCGCGCTGCGGAACTAACCCGCTGCGCATGAG 60

OY 158 TGTCTATCAGAGAGGTGCGGAGAGAGGCTCGGCTATACACGAGCGGATTCAGGCGG 217  
 |||||||  
 DB 61 TGTCTATCAGAGAGGTGCGGAGAGAGGCTCGGCTATACACGAGCGGATTCAGGCGG 120  
 |||||||  
 OY 218 CAGCGCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 277  
 |||||||  
 DB 121 CCGGTGCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180  
 |||||||  
 OY 278 AGCTCAAGAAAGCATATGCGGCGGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 337  
 |||||||  
 DB 181 AGCTCAAGAAAGCATATGCGGCGGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 |||||||  
 OY 338 CGTCTTTCATTTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 397  
 |||||||  
 DB 241 CCTACTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
 |||||||  
 OY 398 CGTCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 457  
 |||||||  
 DB 301 GGTCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360  
 |||||||  
 OY 458 TGAGCGAAAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 517  
 |||||||  
 DB 361 TGAGCGAAAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420  
 |||||||  
 OY 518 AAGGGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 577  
 |||||||  
 DB 421 AAGGGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
 |||||||  
 OY 578 TGATCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 637  
 |||||||  
 DB 481 TGATCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540  
 |||||||  
 OY 638 CCGTTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 697  
 |||||||  
 DB 541 TGTCTATCAGAGAGGTGCGGAGAGAGGCTCGGCTATACACGAGCGGATTCAGGCGG 600  
 |||||||  
 OY 698 GCGGATTCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 757  
 |||||||  
 DB 601 GTCGCTGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
 |||||||  
 OY 758 ACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 779  
 |||||||  
 DB 661 ACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 682  
 |||||||

# RESULT 5

N91423  
 ID N91423 standard; DNA: 1125 BP.

AC N91423.  
 DT 01-FEB-1991 (first entry)  
 DE Sequence of heat-resistant alanine dehydrogenase (AH) gene with mol. wt.  
 DE 2 MD or less  
 KW Enzyme: ds.  
 OS Bacillus stearothermophilus IFO 12550.  
 PN J01043194-A.  
 PD 15-FEB-1989.  
 PF 10-AUG-1987; 200524.  
 PR 10-AUG-1987; JP-200524.  
 PA (NIRA) Unilika K.  
 DR WPI: 89-096096/13.  
 PT Recombinant plasmid for transforming Escherichia coli -  
 PT obd. by connecting heat resistant alanine dehydrogenase gene to  
 PT vector plasmid  
 PS Disclosure: Fig 2, p693; 9pp; Japanese.  
 CC A recombinant plasmid contg. heat-resistant alanine dehydrogenase (AH)  
 CC gene with mol. wt. 2 or less MD is claimed. Cells transformed with the  
 CC vector produce high levels of heat-resistant AH. Transformed E. coli  
 CC is useful as a clinical inspection reagent.  
 SQ Sequence 1125 BP; 246 A; 295 C; 367 G; 217 T;

Query Match 22.4%; Score 277.2; DB 1: Length 1125;  
 Best Local Similarity 56.2%; Pred. No. 1.0e-49;



Matches 604; Conservative 0; Mismatches 458; Indels 12; Gaps 4;

```

QY 61 ATGCGCTCGGATTCGACGAGACCAAAACACAGATTCGGGGGCGCATCACCCG 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 ATGAAGATCGGATTCGACCAAAAGAAATCAAAACCGCGTCGATCAGTCG 60
QY 121 GCGGCGCTCGGAGAACTAACCCGCTGTCGCGCATGAGTGTCT---CATCCAGGAGTGC 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GAGGCGCTGATGAGCGCTGCTCAAAAGCGGGGCGATAGGTGTGTATGTGAGACGAAAGC 120
QY 178 GAGAGGCGCTGCTGATACCGAGCGCGATTTCAAGCGCGAG-TCGCAACTGTGTGCG 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GCGCGCTGCTGCGGGTTCCTCGATTCGAGATGAAAAAGCGGGGAGCTGATCGTTC 180
QY 238 ACCGCGACCGAGGTGTGGCGGAGCTGATTTATTCGTCAGATGCAAGAAAGCATGAGC 297
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 CGAACGTGAGAGATGCTTGGACGCGCGAGATGTTGAAAAGTAAAGAGCCCTGGCT 240
QY 298 GCGGAATACGCGCGCTGCGACAGCGGCGAGATCTTTCACGTTTTCATTTGCGCGCG 357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 CGAGAGTTCCGCTATTTCGCGCGCGAGATTTGTTAGTTTTCATTTAGCGCGCG 300
QY 358 TCAGTGTCTTCACCGAGTGTGTGATTCGCGGACACGTCATTTGCTACGAGACC 417
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GCGGACGCGCTACGAAAGCGCGTGTGAGCAAAAGTGTGCGCATGCTTACGAGACG 360
QY 418 GTCCAGACCGCGCGAGCGGCGACTACCCCTGCTTCCCGGATGAGCAAGTCCGCGTGA 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 GTGAGCTGCGGAAAGCGGTGCTGCTGCACTGTGAACGCGATGAGTGAAGTCCGCGCGC 420
QY 478 CTGCGCGCGCGAGTTGGCGCTTACCACTGATGCGAACCAGGCGCGCGGTGTGCTG 537
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 ATGTCGCGGAGTGTGCGGCGCGCACTTCTTCAGAAAGCGCGGCGGAAAGGCAATTTG 480
QY 538 ATGGCGCGGCGCGCGCGCGGTGGAACCGCGCGAGCGTGTGAGTGTGCGCGCGCGCGCC 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 CTGGCGCGCGCGCGCGCGCGGTGGAACCGCGCGAGTGTGAGTGTGCGCGCGCGCGCC 540
QY 598 GCGTACACCGCGCGCGCGCGCGATGCGCAACGCGATGGCGCGCGCGTTCAGACATC 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 GCGAGCAACCGCGCGCGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 600
QY 658 AACATCGCAAACTTCGCGCACTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGTAC 717
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 601 AACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
QY 718 TCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 777
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 TCAGACTGTGACATATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
QY 778 CTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 837
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 TTGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 777
QY 838 CCAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 894
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 778 CCGGATGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 837
QY 895 GAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 954
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 838 CGAGCGTGTGAGCGAGCGATGCGCATGCGTCAAGCGCGCGTGTGTGTGTGTGTGTGT 897
QY 955 GCGAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1044
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 898 GCGAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 954
QY 1015 CCGTATGTGCTGAGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 955 CCGTATGTGCTGAGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1014
QY 1075 CTAGCGAAAGGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1015 CTGTATTAAGGATGACGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1068

```

```

RESULT 6
V52155/c
ID V52155 standard; DNA: 28171 BP.
AC V52155;
DT 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:22.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KM computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
PN M09818931-A2.
PD 07-MAY-1998.
PF 30-OCT-1997; U19588.
PR 31-OCT-1996; US-029960.
PA (HUMAN) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA;
DR WPI; 98-272225/24.
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae.
PS Claim 1: Page 273-289; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridize to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical.
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 28171 BP; 8507 A; 5080 C; 6440 G; 8144 T;

```

```

Query Match 17 2%; Score 212.4; DB 1; Length 28171;
Best Local Similarity 53.9%; Pred. No. 6.3e-36;
Matches 533; Conservative 0; Mismatches 436; Indels 19; Gaps 4;
QY 61 ATGCGCTCGGATTCGACGAGACCAAAACACAGATTCGGGGGCGCATCACCCG 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21711 ATGTTAATCGGAATCCCAAAAGAAATTAATAAGCAAAACCGTCCGCGCACACCT 21652
QY 121 GCGGCGCTCGGAGAACTAACCCGCTGTCGCGATGAGTGTCTCATCCAGGAGTGC 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21651 GCAAGTGTTCATACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 21592
QY 181 GAGGCGCTCGGATTCACGAGCGCGGATTCAGGCGCGCGCGCGCGCGCGCGCGCGCG 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21591 CTGGTTCCTGCTTACTATGATGCTGACTATCAAAAGCAAGAGTGTGATGTGCTACT 21532
QY 241 GCGGACGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21531 GCTGTGTAAGCTTGGCG---AGCAGAGTGTGTGTGAAAGTAAAGTATCTTAAGTCT 21475
QY 301 GAATAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21474 GAATAGCGTGTACTTGTGCGCGAGGATCTTCTCTTCACTTGTGACATGAGCGCGTCT 21415

```



```

QY 361 CGTGTTCACCGATGCGTGTGGATTCGGGACACGTCATTCGATCGAGACCGTC 420
DB 21414 CCAGAAATTTACGATGCTATGTT-----AACACCAAAAACACTGAACCTGT 21367
QY 421 CAGACCGCGAGCGGACGACACCGCTGTCGCCCGATGAGCGAGTCGCCGGCGACTC 480
DB 21366 CGTACACATCAAGGACACGATCGCTCTGCTTCTATAGAGTGAGTGACGATGATG 21307
QY 481 GCGCCCGAGGTGGCGCTTACACCTGATCGAACCCAGGGGGCGCGGTGCTGATG 540
DB 21306 GCTGTTCATTAATTCGAGAGTCACTTCTACTAGACAGCTGTGGCTGTGTTCTACTT 21247
QY 541 GCGGGGTGCGCGCGCTGACACCGCGGAGCTGTGTATTCGGCGCGCGCGCGCGCG 600
DB 21246 GGTGTGTACAGGTGTTCACAAAGAAAGTAACTCATCGTGTGTGTGTGTGTGT 21187
QY 601 TACAAAGCAGCGCGCATCGGCAAGGATGGCGCGACCTTACGTTAGATGATCAAC 660
DB 21186 ACACATGCTGCGCGCATCGCGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 21127
QY 661 ATGCAAACTTGGCACTCGACCGCGAGTTCGCGCGGATCCACACTCGCTACTCA 720
DB 21126 TCCAAAGGCTCTCTAGTTCTAGAAAGATCTTGGAGTCA/ATTCAAACTTATGTCT 21067
QY 721 TCGGCTTACGAGCTCGAGGCGCGCTCAACGCGCGACCTGTGATGGCGCGCTCTG 780
DB 21066 AATTCAATTCACATTAAGCAAGTGTAGAGATCTGATGTGTGTGTGTGTGTGTGT 21007
QY 781 GTGCGAGGCGCGCAAGCGCAATTAATGTCTGCAATTCAGTGTGCGCATATGAAACA 840
DB 21006 ATCCCTGTGTGCAAAAGCAGCGGATGTGTGACAGATGATGATGCAAAAGTGTGCA 20947
QY 841 GGTGCGTACTGTGTGATTAATGATGATGACGAGGCGCGCTGTGTGCA--AGGCTACGA 897
DB 20946 GGTCTCTTA--TCGTTGACGTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 20888
QY 898 CCGACCACTTACGACGCGCGAGTTCGCGGTGCGGACGACGAGCTTCTGTGTGTGTGT 957
DB 20887 GTGCAACGCGACGATGAAACCGCTTATGAAACACGCTGTCTTCCACTATGCGCTGTCC 20828
QY 958 AACATGCGCGCGCTGCTGTGCGCAAGACGCTGACGCTGACGCAACGCGGACGATGCG 1017
DB 20827 AATATCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 20768
QY 1018 TAATGCTTCAAGCTTGTGCGGACGATGCT 1045
DB 20767 TATATCGAAGCTTGTGCTGTGCAAGGAT 20740

RESULT 7
V74442/c
ID V74442 standard; DNA; 9280 BP.
AC V74442;
DE 16-MAR-1999 (first entry)
KW Staphylococcus aureus confg SEQ ID #131.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
OS Staphylococcus aureus.
FH Key location/Qualifiers
misc_feature 841..900
FT /tag- a
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 2641..2700
FT /tag- b
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"

```

```

FT misc_feature 4441..4500
FT /tag- c
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 6241..6300
FT /tag- d
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 8041..8100
FT /tag- e
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT BP-786519-A2.
FT 30-JUL-1997. 100117.
FT 07-JAN-1997; 100117.
FT 05-JAN-1996; US-009861.
FT (HUMA-) HUMAN GENOME SCI INC.
FT Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
FT PI Rosen CA;
FT WPI: 97-374922/35.
FT DR
FT WPI: 97-374922/35.
FT PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
FT stored on computer readable medium and used in the production of
FT anti-S.aureus vaccines
FT PS Claim 1: Page 710-715: 3271pp: English.
FT CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
FT of the invention. The DNA sequences are recorded on a computer readable
FT medium, preferably selected from a floppy or hard disk, random access
FT memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
FT CC the S.aureus DNA sequences allows putative functions to be assigned so
FT CC that protein-encoding or regulatory regions of commercial, therapeutic or
FT CC industrial importance can be obtained. Specifically, sequences which are
FT CC likely to encode antigens have been identified and these polypeptides can
FT CC be used in a vaccine composition against S.aureus infection. The
FT CC polypeptides can also be used in a kit for the immunodetection of
FT CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
FT CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
FT CC skin and surgical wound infections, scalded skin syndrome, toxic shock
FT CC syndrome, etc. Organisms transformed with the DNA sequences can be used
FT CC for recombinant production of the polypeptides. The new DNA sequences
FT CC (and their fragments) are useful as primers or probes for isolating
FT CC homologues of any of the S.aureus DNA sequences contained on the
FT CC computer readable medium.
FT SQ Sequence 9280 BP; 2958 A; 1649 C; 1309 G; 3060 T;

Query Match 14.4%; Score 178.4; DB 1; Length 9280;
Best Local Similarity 50.6%; Pred. No. 6,7e-29;
Matches 484; Conservative 0; Mismatches 466; Indels 6; Gaps 2;

QY 151 CATGAGGTGCTCATCCAGCAGTGGCGGAGGCTGCGTATCCACGAGGATTC 210
DB 9280 CATATGTTTAACTTAAGAACACACTGATATTCATTTTAAAGAGGCTTGA 9104
QY 211 AAGCGCGAGCGCGGACGATGCTGCGGACCGCGGATGCGGCGGCGGCGGCGGCGG 270
DB 9220 AAAGAAAGCAGGTGCTGATGATTTGCTGCAACAAAGCAAAAGTTGG--GATGTGATATG 9164
QY 271 TTGCTCAAGGTCAAAAGACGATAGCGGATAGCGCGGCTGCGGACGCGGCGATC 330
DB 9163 GTATATTAAGTAAAGAACACACTGATATTCATTTTAAAGAGGCTTGA 9104
QY 331 TTGCTCAAGGTGCTGATTTGGCGGCTGACGCTGCTTGAACGCGATCGTTGTGATTC 390
DB 9103 TATATCACTTATTTTATTTAGCAATGAAGAAATTAACACACTTGTATGATAGA 9044
QY 391 GGCACCGCGTCAATGCTGTGAGAGCGGTCACAGCGGCGGCGGCGGCGGCGGCGGCT 450

```

```

Db 9043 AAGTAATAGTATTGATATGATAGACTGTGCACTTACAGACCGATCTTACCATTTGTA 8984
QY 451 GCCCGATAGAGGATGCGCGGTGACCTGCGCCAGGTTGGCGCTTACACCGGATG 510
Db 8983 TCACCAATAGTAGAGTACAGAGAAATGTATGCTCAAGTTGGCGCAGATTTCTACAA 8924
QY 511 CCACCCCAAGGGGGCGCGGTGTGTGATGGCGGGGTGCGCGGTGCGACCGCGAC 570
Db 8923 AACTTAAATGATGATAGGATTTCTACTAGTGTGTGCTCCAGGATTAAGGTTAA 8864
QY 571 GTGTGTGTATGGCGCGCGGACCGCGCTTACACGACCGCGGATCGGCAACGGCATG 630
Db 8863 GATCAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8804
QY 631 GCGGCGACCGTACGCTTACATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 690
Db 8803 GGTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8744
QY 691 TTCTGCGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 750
Db 8743 TTCTGCGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8684
QY 751 CGTCCGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 810
Db 8683 CAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8624
QY 811 TCGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 870
Db 8623 ACAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8564
QY 871 CAGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 927
Db 8563 CAAAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8504
QY 928 GTGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 987
Db 8503 AAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8444
QY 988 ACCTACGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1047
Db 8443 ACCTACGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8384
QY 1048 CGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1103
Db 8383 AGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8328

RESULT 8
ID T67971 standard; DNA; 1074 BP.
AC T67971:
DE 15-JUL-1997 (first entry)
KM H. pylori membrane protein ORF 05sep2051orf1.
KM Vaccine prevention; treatment; infection; identification;
KM binding compound; bacterium; life cycle; activator; bacteria;
KM inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KM membrane; amino acid; metabolism; ds.
OS Helicobacter pylori.
FH key Location/Qualifiers
FT cds 1..1074
FT /tag= a /note= "no stop codon given"
PD W09640893-A1.
PD 19-DEC-1996.
PD 06-JUN-1996: U09122.
PR 07-JUN-1995: US-487032.
PR 01-APR-1996: US-630405.
PR (ASTR) ASTRA AB.
PR Berglindh OT, Smith D, Mellgaard BL.
DR P-PSDB: W20718.
PT Helicobacter pylori nucleic acid sequences and related
polyptide(s) - useful for vaccines to treat or prevent H. pylori

```

```

PT infection, and to detect Helicobacter
PS Claim 27: Page 819: 1481pp: English.
CC The present sequence encodes a Helicobacter pylori membrane
CC protein likely to contain four membrane spanning regions.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the
CC bacterial DNA. The sequences were analysed for ORF of at least 180
CC nucleotides, and the predicted coding regions defined by computer
CC evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences
CC of interest, particular regions can be isolated from H. pylori by
CC PCR amplification for recombinant polypeptide production, e.g. in
CC E. coli hosts.
SQ Sequence 1074 BP; 302 A; 189 C; 297 G; 286 T;

```

```

Query Match 7.7%; Score 94.8; DB 1; Length 1074;
Best Local Similarity 46.1%; Pred. No. 1.36-11;
Matches 370; Conservative 0; Mismatches 417; Indels 15; Gaps 1;

```

```

QY 265 GATTATTGCTCAGGATCAAGACCATAGGCGGGAATGCGCGCTGCGACACGG 324
Db 133 GATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 192
QY 325 CAGATCTGTTGCTCAGGATCAAGACCATAGGCGGGAATGCGCGCTGCGACACCATG 384
Db 193 GCGACCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 252
QY 385 GATTCCGCGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 444
Db 253 AATAAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 312
QY 445 CTGCTGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 496
Db 313 ATTTAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 372
QY 496 -----GCTTACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 549
Db 373 TTACTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 432
QY 550 CCCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 609
Db 433 TCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 492
QY 610 GCCCGATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 669
Db 493 GCGAAGCTTTAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 552
QY 670 CTTCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 729
Db 553 TTACAAACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 612
QY 730 GAGCTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 789
Db 613 AATATCATTTCAAGGCTTAAACGCGGATGATGATGATGATGATGATGATGATGATGAT 672
QY 790 GCCAAGGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 849
Db 673 AGCCAAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 732
QY 850 CTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 909
Db 733 GTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 792
QY 910 GACCAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 969
Db 793 TCTAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 852

```





RESULT 11  
Q21833/c  
ID Q21833 standard; DNA: 390 BP.  
AC Q21833;  
DT 08-JUN-1992 (first entry)  
DE Randomising oligonucleotide used in SPERT mRNA prep.  
KW Systematic polypeptide evolution by reverse translation; SPERT;  
ligand binding; ss.  
OS Synthetic.  
PN WO9202536-A.  
PD 20-FEB-1992.  
PF 01-AUG-1991; U05463.  
PR 02-AUG-1990; US-561968.  
PA (COLS) UNIV OF COLORADO.  
PI Gold L, Tuerc C;  
DR WPI: 92-08008/10.  
PT New method of systematic polypeptide evolution by reverse translation - by linking each polypeptide in sample mixt. to individualised mRNA allowing further synthesis of selected polypeptide(s)  
PS Example: Page 55: 102pp; English.  
CC The sequence is that of an example randomising oligonucleotide which is used in the prepn. of mRNA encoding candidate polypeptides for the method of systematic polypeptide evolution by reverse translation (SPERT). The method provides a rapid way of isolating and identifying polypeptide ligands which bind to target mols. The polypeptide ligands can be used in e.g. assay methods, diagnostic procedures, cell sorting, as activators or inhibitors of target mol. function, as probes, as sequestering agents, drug delivery vehicles, modifiers of hormone action and as catalysts. See also Q21830-Q21832.  
CC Sequence 390 BP; 125 A; 126 C; 133 G; 6 T;  
SQ

Query Match 4.1%; Score 50.8; DB 1; Length 390;  
Best Local Similarity 46.2%; Pred. No. 0.016;  
Matches 169; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

QY 426 CGCCGACGGCGACACTACCCCTGCTCCCGGATGAGCGAAGTCCGCGTGCATCGCCGC 485  
DB 366 CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 307  
QY 486 CCAAGTTGGGGCTTACCTGATGCGAACCAGGGGGCGGGTGTCTGATGGGCGG 545  
DB 306 CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 247  
QY 546 GGTGCGCGGCGGACG 605  
DB 246 CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 187  
QY 606 CGGACCG 665  
DB 186 CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 127  
QY 666 CAAACTTGGGCACTGACG 725  
DB 126 CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 67  
QY 726 CTACGAGCTCGAAGGCG 785  
DB 66 CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7  
QY 786 AGGCGC 791  
DB 6 TGGCCC 1

RESULT 12  
Q36859/c  
ID Q36859 standard; DNA: 390 BP.  
AC Q36859;  
DT 22-JUN-1993 (first entry)

DE PCR primer for 5' fixed sequence contg. T7 promoter and RBS.  
KW Systematic peptide evolution by reverse translation; SPERT; ligand;  
specific; inhibitors; probes; assay; cell sorting; ss.  
OS Synthetic.  
PN WO9303172-A.  
PD 18-FEB-1993.  
PF 31-JAN-1992; U00801.  
PR 01-AUG-1991; US-739055.  
PA (UYRE-) UNIV RES CORP.  
PI Gold L, Pribnow D, Smith JD, Tuerc C;  
DR WPI: 93-076529/09.  
PT Systematic polypeptide evolution by reverse translation - used for prodn. of polypeptide ligand specific for desired target molecule.  
PS Example 1: Page 84; 98pp; English.  
CC SPERT is used to select novel polypeptides that bind the antibody of the epitope commonly recognised by the antisera from autoimmune mice which are the fl progeny of a cross of NZB and NZW parents (Portanova et al.; J. Immunol. 144, 4633, 1990). The known epitope consists of ca. 10 amino acids at the N-terminus of the histone H2B protein. To make mRNA encoding candidate polypeptides a 5' fixed sequence composed of a T7 promoter sequence and a ribosome binding site which is recognised by both prokaryotic and eukaryotic ribosomes, terminating in a restriction endonuclease site is synthesised and cloned using a number of oligonucleotides (example shown). A 3' fixed sequence is placed into a restriction site to provide an mRNA encoding the C-terminal trailer sequence of ca. 100 nucleotides lacking stop codons. In addition, a 3' primer annealing site is provided so that cDNA synthesis can be accomplished on the mRNA recovered from partitioned CC ribosome complexes. See also Q36845-63.  
CC Sequence 390 BP; 125 A; 126 C; 133 G; 6 T;  
SQ

Query Match 4.1%; Score 50.8; DB 1; Length 390;  
Best Local Similarity 46.2%; Pred. No. 0.016;  
Matches 169; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

QY 426 CGCCGACGGCGACACTACCCCTGCTCCCGGATGAGCGAAGTCCGCGTGCATCGCCGC 485  
DB 366 CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 307  
QY 486 CCAAGTTGGGGCTTACCTGATGCGAACCAGGGGGCGGGTGTCTGATGGGCGG 545  
DB 306 CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 247  
QY 546 GGTGCGCGGCGGACG 605  
DB 246 CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 187  
QY 606 CGGACCG 665  
DB 186 CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 127  
QY 666 CAAACTTGGGCACTGACG 725  
DB 126 CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 67  
QY 726 CTACGAGCTCGAAGGCG 785  
DB 66 CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7  
QY 786 AGGCGC 791  
DB 6 TGGCCC 1

RESULT 13  
X34205  
ID X34205 standard; DNA: 1413 BP.  
AC X34205;  
DT 06-JUL-1999 (first entry)  
DE Mycobacterium species nucleic acid sequence 415.  
KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;



